```
OM protein - protein search, using sw model

                            GenCore version 5.1.3
                            Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:              January 7, 2003, 16:25:31 ; Search time 81 Seconds
                            (without alignments)
                            1778.110 Million cell updates/sec

Title:               US-09-809-638-2
Perfect score:       3720
Sequence:            1 MTSLWREILLESLLGCVSWS.........EGHNYENNHHFHMNTPKYFL 699

Scoring table:       BLOSUM62
                     Gapop 10.0 , Gapext 0.5
Seq:ed:              671580 seqs, 20604715 residues

Total number of hits satisfying chosen parameters:    671580

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :           SPTREMBL_21:*
                     1: sp_archea:*
                     2: sp_bacteria:*
                     3: sp_fungi:*
                     4: sp_human:*
                     5: sp_invertebrate:*
                     6: sp_mammal:*
                     7: sp_mhc:*
                     8: sp_organelle:*
                     9: sp_phage:*
                     10: sp_plant:*
                     11: sp_rodent:*
                     12: sp_virus:*
                     13: sp_vertebrate:*
                     14: sp_unclassified:*
                     15: sp_rvirus:*.
                     16: sp_bacteriap:*
                     17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match Length	DB ID	Description
1	3713	99.8	699	4	Q9H720	Q9H720 homo sapien
2	3125	84.0	699	11	Q9YL7	Q91yl7 mus musculu
3	671.5	18.1	714	5	Q8T1C4	Q8t1c4 dictyosteli
4	639	17.2	425	3	Q9P6M4	Q9p6m4 schizosacch
5	625.5	16.8	401	3	P78781	P78781 schizosacch
6	256	6.9	688	3	Q9HD22	Q9hd22 schizosacch
7	141.5	3.8	279	16	Q9S1Z5	Q9s1z5 streptomyce
8	133	3.6	466	16	Q8Z4F6	Q8z4f6 salmonella
9	132	3.5	583	8	Q950T2	Q950t2 hyaloraphid
10	129	3.5	659	6	Q8SPK7	Q8spk7 sus scrofa
11	128.5	3.5	396	2	P95556	P95556 pseudomonas
12	128	3.4	466	16	Q82MM0	Q82mm0 salmonella
13	124.5	3.3	487	16	Q82R15	Q82r15 salmonella
14	124	3.3	609	8	O79569	O79569 mustelus ma
15	123.5	3.3	269	11	Q91VY8	Q91vy8 mus musculu
16	122	3.3	636	6	Q9GJX6	Q9gjx6 sus scrofa

```
17  122      3.3  121   6   Q9GJX7   Q9gjx7 sus scrofa
18  121.5    3.3  766  17   Q9UYP5   Q9uyp5 pyrococcus
19  121      3.3  591   8   Q36149   Q36149 trachenys s
20  121      3.2  605   8   Q9B6Y4   Q9b6y4 casuarius b
21  119      3.2  427  16   O53515   O53515 mycobacteri
22  119      3.2  612   8   Q9G03    Q9gq3 diplophos t
23  118.5    3.2  449  16   Q9RL05   Q9rl05 streptomyce
24  118.5    3.2  605   8   Q95720   Q95720 casuarius c
25  118.5    3.2  613   8   Q94SY9   Q94sy9 creninugil
26  118.5    3.2  694  16   Q91J79   Q91j79 streptomyce
27  118      3.2  431  16   Q82C80   Q82c80 yersinia pe
28  117.5    3.2  501  16   Q82833   Q82833 salmonella
29  117.5    3.2  583   5   O17069   O17069 caenorhabdi
30  117.5    3.2  559  16   Q82B84   Q82b84 salmonella
31  117      3.1  570   8   Q9B624   Q9b624 apteryx man
32  116.5    3.1  613   8   Q9B510   Q9b510 tetrodontop
33  116.5    3.1  557  16   Q94SZ9   Q94sz9 mugil cepha
34  116.5    3.1  605   8   Q8X9F8   Q8x9f8 escherichia
35  116      3.1  605   8   Q9B6W0   Q9b6w0 dinornis gi
36  116      3.1  605   8   Q9B6U8   Q9b6u8 dromaius no
37  116      3.1  605   8   Q958A0   Q958a0 dromaius no
38  116      3.1  958  16   Q82FM4   Q82fm4 yersinia pe
39  115.5    3.1  590  16   Q96XC0   Q96xc0 sulfolobus
40  115      3.1  317  16   Q97K74   Q97k74 clostridium
41  115      3.1  603   8   Q9MR50   Q9mr50 ciconia cic
42  115      3.1  603   8   Q958D4   Q958d4 ciconia cic
43  114.5    3.1  603   8   Q9GA20   Q9ga20 tinamus maj
44  114.5    3.1  829  10   Q9LMJ1   Q9lmj1 arabidopsis
45  114      3.1  395   2   Q9Z404   Q9z404 pseudomonas
```

ALIGNMENTS

```
RESULT 1
Q9H720

ID   Q9H720     PRELIMINARY;    PRT;   699 AA.
AC   Q9H720;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CDNA: FLJ21511 fis, clone COL05748.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON;
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA   Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK025164; BAB15080.1; -.
SQ   SEQUENCE 699 AA;  78565 MW;  2F839563189A0523 CRC64;
```

Query Match   99.8%;
Best Local Similarity   99.8%;   Score 3713; DB 4; Length 699;
Matches 698; Conservative 1; Mismatches 0; Indels 0; Gaps 0; Pred. No.3.6e-263;

```
Qy    1 MTSLWREILLESLLGCVSWSLYHDLGPMIYFPLQTLELTGLEGFSIAFLSPIFLTIPF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MTSLWREILLESLLGCVSWSLYHDLGPMIYFPLQTLELTGLEGFSIAFLSPIFLTIPF 60

Qy   61 WKLVNKKWMLTLLRITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 WKLVNKKWMLTLLRITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120

Qy  121 LRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNRKVILTLSAIATLDRIGTDGDCSKPEEK 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 LRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNRKVILTLSAIATLDRIGTDGDCSKPEEK 180
```

RESULT 2

```
Qy  181 KTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGDPNPFGGA 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 KTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGDPNPFGGA 240

Qy  241 VLLCLASGLMLPSCLWFRCTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTASMPQ 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VLLCLASGLMLPSCLWFRCTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTASMPQ 300

Qy  301 TLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLII 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLII 360

Qy  361 GLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLRHKAYERKLG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLRHKAYERKLG 420

Qy  421 KVAPTKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFIITLLESDASKPYMGNNDLT 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 KVAPTKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFIITLLESDASKPYMGNNDLT 480

Qy  481 MWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAILTLTVNISGK 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 MWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAILTLTVNISGK 540

Qy  541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNV 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNV 600

Qy  601 KDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 KDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660

Qy  661 DHREVSEKIHFNPRFGSYKEGHNYENNHHFHMNTPKYEL 699
        |||||||||||||||||||||||||||||||||||||||
Db  661 DHREVSEKIHFNPRFGSYKEGHNYENNHHFHMNTPKYEL 699
```

```
ID   Q91YL7       PRELIMINARY;   PRT;   699 AA.
AC   Q91YL7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Similar to hypothetical protein FLJ21511.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RL   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC016523; AAH16523.1; -.
DR   InterPro; IPR001092; HLH_basic.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE 699 AA; 78179 MW; 4F41975D6570D5F8 CRC64;
```

Query Match   84.0%; Score 3125; DB 11; Length 699;
Best Local Similarity   82.8%;
Matches 579; Conservative   41; Mismatches   79; Indels   0; Gaps   0;

```
Qy    1 MTSLWRETLLESLLGCVSWSLYHDLGPMIYYFPPLQTLELTGLEGFSIAFLSPIFLTIPF  60
        |||||| || || ||| |||||||| ||||||| ||||||||||| |||||||||||| |
Db    1 MPGLWRAIALETLLGYVSWSLYHGLSPMIYYPPLQTLELTGLEFFCVAFLSPILLTIPPL  60

Qy   61 WKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120
        ||||||| ||:|||| |: ||||| |||||||||||||||||||||| ||||||| ||||
Db   61 WKLVNKKWTLSLLRIVTVGSIASFEAPNAKLRLMVLALGVSSSLIVQTVTWWSGSGLQRY 120

Qy  121 LRIWGFILGQIVLVLRIWYTSLNPIWSYQMSNKVILTLSAIATLDRIGTDGDCSKPEEK 180
        |:|||||| ||:|||||||||||||||||||| |||||||:|||||||||||||||:| |
```

RESULT 3

```
Db  121 LKIWGFILGHVLLLVLRIWYTSLNPIWSYQMSNRVILTLSAVAVLDRIGTDGDYRNPEGK 180
        | |||||| || :|:||||||||||||||||||:|||||||:|:|||||||||  ||:||
Qy  181 KTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGA 240
        |  |||||   :  :||  ||||||:|||||:||  |:|||||||||||||| |||||||
Db  181 KPREVATGRYSLSSWLLPGAAFGSLLFLTHWIFGGVSIVSRWAVSGHPHPGDPNPFGGA 240

Qy  241 VLLCLASGLMLPSCLWFRCTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTASMPQ 300
        ||| : ||||| |  | || ||:|| |||:|| |||||:||||||||||: |:| ||:|
Db  241 VLLGFSSGLMLSGSSWLHDAGLAWWMTGAASAMGLLYLRHTWAAAVSGCVLAVFTGSMWPQ 300

Qy  301 TLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLII 360
        |:|||||||||||||||||:||||||||||||||||||||||||||||||||||||::||
Db  301 TIGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLII 360

Qy  361 GLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLRHKAYERKLG 420
        || ||||||||||||||||||||::||:||||||||||:|||||::|:| | |:|:|
Db  361 GLSMLFGPKKNLDFLLQTKNSPKFLLRCSEKYMKLILWLFVGVGLLGLGLRHRTYERQLG 420

Qy  421 KVAPTKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFIITLLESDASKPYMGNNDLT 480
        |:|   :| | |  |:| || ||||||||||||||||:||:|||||||||||||:||||||
Db  421 RGAPATVVSAAIWPFRFGYDNEGWPNLERSAQLLKETGADFIITLLESDASKPYIGNNDLT 480

Qy  481 MWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAILTLTVNISGK 540
        |||||||||||||||||:|:||||:|:|||||||||||||: |: |||| |||:|||||:
Db  481 MWLGEKLGFYTDFGPSTRDHTWGIMVLSRYPIVRSEHHLLPSPEGEIAPAITMTVNVSNR 540

Qy  541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNV 600
        |||||||||||||||||||||||||||:|:|||||||||||:| ||||||||:||||||
Db  541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKNCSNQVIFLGYITSEPGSRDYIQLTKHGNV 600

Qy  601 KDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660
        ||||| || ||||||||||||||||||||||||||||||||||||||| |||||||||||
Db  601 KDIDSSGGDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPANYRDNQKVVI 660

Qy  661 DHREVSEKIHFNPRFGSYKEGHNYENNHHFHMNTPKYEL 699
        |||||| |||||||||||||||| |:|:|||||||||||:
Db  661 DHRGVPKNIHFNPRRFGSYKEGHNVENTHFHFHMNTPKYEV 699
```

```
ID   Q8T1C4       PRELIMINARY;   PRT;   714 AA.
AC   Q8T1C4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 79.7 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., Abril J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116960; AAM08475.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE 714 AA; 79745 MW; 82F773B4CC75C349 CRC64;
```

Query Match   18.1%; Score 671.5; DB 5; Length 714;
Best Local Similarity   27.1%;
Matches 188; Conservative 147; Mismatches 281; Indels   79; Gaps   22;

```
Qy    8 ILLESLLGCVSWSLYHDLGPMIYYFPPLQTLELTGLEGFSIAFLSPIFLTTPFWKLVNK-  66
        :|| : | |:||||: ::|||||| |  || | |||| :  | |:| :|:|:|: |||:
Db   25 LLADGILAFVNWSLMVEFISVIWFFPMGFTGWEAALLAIFTPPIFLGIFINRIVSRF     84

Qy   67 ----KWMLTLLRIITI------GSIAS------FQAPNAKLRLMVLALGVSSSLIVQAVT 110
           :|||||||||||      |||||      |||||||||||||||||||||||: :
Db   85 NLPIRLVAQLLAFPAIFPWAIDKASIKNDEFGIFSSPHV-TKTFFVSVAVALDWLAQCNS 143
```

```
Qy  111 WWS-GSHLQRYLRIWGFIGQIVLVVLRIWYTSLNPI--WSYQMSNKVLILTLSAIATLDR 167
            |:|:| | |  |::|: |: | :  :| :|   :| :| |:|::| |:  || | |:
Db  144 FANVKSAVRRERGAYAYALAIVLHAIIIRLPYLSINPFMTWSNWIVFGMLILVLAPVIL-- 201

Qy  168 IGTDGDCSKPEEKKTGEV--ATGMASRPNWLLAGGAAFGSLVFLTHWVFGEVSLVSRWAVS 225
            |:|:|:| ||     :|   ||  :| | ||: | |:|:|| :| | |: :| | |  |
Db  202 MGENINAS--DEKKSGAIWSIEHETSSSSTIATGVSFGGIIFFNQLLFSTYGLIPRWV-- 257

Qy  226 GHPHPGDPDNPFGGAVLLCLASGLMLPS---CLMFRG------TGLIWWVTGTASAAG 274
            | | |  |: || || | |    :|: :    |:|:         || ||: |:| |
Db  258 ------GLNPPPYGLFVILGMMAGVVISKKHLVTSRNFFIFALILSAIFGWCSGTVP--- 309

Qy  275 LLYLHTWAAAVSGCVFAIFTASMMPQTLGHLLINSGTNPGKTMTIAMIFYLLEIFFCAWCT 334
            | || || |:|  |:|  | |::|||| ||||||||||||||| |||:|||||::|:|::
Db  310 --FIVGFIGLISASLLGAYCNSLMIVLIEKVSNI-EKLQOLFTSAMLTYTVLLFWAIYVV 366

Qy  335 AFKFVP---GGVYARERSDVLLGTMMLLIGLNMLF------GPKKNLDLLLQTKNSSKVL 385
           | | | |      | |  :  || :|:||::|: |        |||||:|::|:: |:|
Db  367 SYKFVPWWLGSTLLRERNQTLIACVLAQGIAFVYKSLRSSGGSSSSIKREKSNKDQPS 426

Qy  386 FRKSEKYMKLF-LWLLVGVGLLGLGLRHKA------YERKL-----GKVAPTKEVSAA 431
            || |  | ::  || |  ::||:|:|:|          ||         |||| | |
Db  427 YPNKDVFNILFGLLVLLLLFSVNRAITHPTDSSIAGHHYTRDLTATSVTNTAPT-DIKSM 485

Qy  432 IWPFRFGYDNEGWSSLERSAHLLNETGADFITILESDASKPYMGCNNDLIPMWLGEKLGFYT 491
            |||:| | ||: |||||||||||||||| |||||| ||||||| ||||:|:| | | |||
Db  486 IWTIHFGYDNFGRNSFPNVTKVIKDHGANIIGLLESDLSRVMTSNRDLVEWIATELHMYS 545

Qy  492 DFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGGEIAPAILTLTVNISGKLVDFVVTHFGN 551
            | | || || || :|||:|||||||||||| || |:||||: | |||: || |||:|| |
Db  546 DFGPEPSRNTWGCALLTIFPIISSDRVILPSPEGELACLIDAVLLIDETPVNVIVTHFGN 605

Qy  552 HEDDLDRKLQAIAVSKLLKSSSNQ--VIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDHDR 610
            | || |||||||||| ||: |      || |||||| ||||||||| |||||| | | |
Db  606 TEDVLDRKLQTEGAAAIVKKNADMPIFLSYITTKVNTENYNTLRASGL---DDTTNEGR 662

Qy  611 WCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRI 645
            :| ||:|||||||||:| |||||| |::::|| |:
Db  663 YCQIIFYKNLQLNKFQRSFSFDISDTEQVASFSV 697
```

RESULT 4

Q9P6M4

Query Match   17.2%; Score 639; DB 3; Length 425;
Best Local Similarity   34.2%; Pred. No. 1.3e-38;
Matches 152; Conservative 70; Mismatches 178; Indels 44; Gaps 10;

```
Qy  276 LYL-HTWAAAVSGCVFAIFT-----ASMMPQTLGHLLINSGTNPGKTMTIAMIFYLLEIFF 329
            |:| | | | | :||::||          |:|||||||||||||||| |:|::|||||||
```

```
Db    4 MYINHGWCSYLGGLIFTSYVLIYSPASI--------RISSFYSPAKWGGAFLVYILYSLA 56
            |  | |  | |  |::| |:  :     |         |:|| ||||| |||:|:||:|
Qy  330 CAWCTAFKFVPGGVYARERSDVLLGTMMLLIGLNMLF--------GPKKNLDLLLQTK 379

Db   57 HVWVVAYEFVPGGPILRERTSYIL----IFIGWNLAALVPAYSGESKEPNKADSSVVDIK 112
           ::|: | | |||  | | |: |          :  |  ||:| | | |: | |  | |
Qy  380 NSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYERKLGKVAP---TKEVSAAIWPF 435

Db  113 QSDSSYRRRSFKKS-----LLTGFCLALMALKFAIQNMPPYDYTPYHPNEKLFTAGIWTI 167
            | |:|   | |:       |||| | |:| | :: : ||||:|| || || | |:|:|
Qy  436 RFGYDNEGWSSLERSAHLLNETGADFITILESDASKPYMGCNNDLIPMWLGEKLGFYTDFGP 495

Db  168 HFGLDNFMYASENRIRDAVRDMELDVFGLLESDTQRLIMGFRDLIQVLAHDLGMYADYGP 227
           ||| ||| | || ||| | | ||||| || | | ||| :| |||  | :| | | |  ||
Qy  496 STRYHTWGIMALSRYPIVKSEHHLLPSPEGGEIAPAILTLTVNISGKLVDFVVTHFGN 555

Db  228 GPDKHHWGAALLSKFPIVNSTHHLLPSPQGELAPAIHATLDVYGELIDVVVSHNGQYESQ 287

Qy  556 LDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDHDRWCEYI 615
            ||||||||:|  |::   ::|:|||||||| ||||||||| |||||| |  | | :|:| *
Db  288 LDRRLQSTELARIMRESPRPLVFLGIVVVSNVGQOEPQTILFTRDTGMLDIEPADYDRNCQYI 347

Qy  616 MYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVIDHREVSEKIHFNPRF 675
            |||||||||||||||| | |::::||| |    | : | : :  | |: |: | |  |
Db  348 FYRGVKRIGYARLRHSTITDTELQTGKFLVTKDL----GRNVRIDKEHVPESHRYPSLF 402

Qy  676 -GSYKEGHNYENNHHHFHMNTPKYF 698
           ||| |  | | | :|:   |||| |
Db  403 EGTGVNGHYHYDNNLVVH--EPWYY 424
```

RESULT 5

P78781

Query Match   16.8%; Score 625.5; DB 3; Length 401;
Best Local Similarity   35.5%; Pred. No. 1.2e-37;
Matches 145; Conservative 63; Mismatches 169; Indels 31; Gaps 7;

```
Qy  306 INSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLLIGLNML 365
            |:|| ||||| |:|::||||||||:| | | | ||||  | | |: ||| :|:|||:|
Db    9 ISSFYSPAKVWGGAFLVYILYSLAHVWVVAYEFVPGGPILRERTSYIL----IFIGWNLA 64

Qy  366 F-------GPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAY 415
                    |||||:|::|::|:| ::|  | | |:|| || | |:||:|:|:|| |
Db   65 ALVPAYSGESKEPNKADSSVVDIKQSDSSYRRRSFKKS-----LLTGFCLALMALKFAIQ 119

Qy  416 ERKLGKVAP----TKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFITILESDASK 471
            | | ||      || | | |||| ||| |||:||||||||||||||| |||||| |||
Db  120 NMPPYDYTPYHPNEKLFTAGIWTIHFGLDNFMYASENRIRDAVRDMELDVFGLLESDTQR 179

Qy  472 PYMGCNNDLIPMWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGGEIAPAI 531
```

```
Db    180 LIMGFRDLTQVLAHDLGMYADYGPGPDKHFWGAALLSKFPIVNSTHHLLPSPQGELAPAI 239
              ||| |:  | |||:::|  |||: ||||| |::|||| || |||:|||||:||:|||||
Qy    532 TLTVNISGKLVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDY 591

Db    240 HAILDVYGELIDVVVSHNGQYESQLDRRLQSTELARIMRESPRPLVFLGYVVSNVGQEPQ 299
              ||:| |:|  ||||||   ||||:     ||| :||:  |:||||||||||  |||| |
Qy    592 LQLTEHGNVKDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTN 651

Db    300 TILTRDTGMLDIEPADYDRWCQYIFYRGYKRIGYARLHRSTITDTELQTGKFLVTKDL-- 357
             :|| |:||| :|::|||||| ||:  ||  :|||:| ||| :|::|| |:|||  ||
Qy    652 YRDNQKVVIDHREVSEKIHFNPRF-GSYKEGHNYENNHHHFHMNTPKYF 698
              ||  |:  ||  | | ||: |    |:|:: :|::|:||:| ||:|
Db    358 ---GRNVRIDKEHVPESHRYPSLFEGTGVNGHYYDNNLVVH--EPWYY 400
```

```
Q9HDZ2
ID   Q9HDZ2;         PRELIMINARY;      PRT;       688 AA.
AC   Q9HDZ2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Hypothetical 77.5 kDa protein (Fragment).
GN   SPAC589.12.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Wood V., Rajandream M.A., Barrell B.G., Aert R., Robben J.,
RA   Weltjens I., Grymonprez B., Volckaert G.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL512496; CAC19769.1; -.
KW   Hypothetical protein.
FT   NON_TER      688
FT   NON_TER      688
SQ   SEQUENCE   688 AA;  77474 MW;  3985EAC23CDA34FF CRC64;

Query Match             6.9%;  Score 256;  DB 3;  Length 688;
Best Local Similarity   22.5%;  Pred. No. 2.6e-10;
Matches     97;  Conservative   77;  Mismatches 197;  Indels  60;  Gaps  11;

Qy     11 ESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEGFSIAPLSPIFLTITPFWKLVNKKWML 70
             |: ::||| ||::::|| |  ::|| ::||: ||| ||| :| : | |:| |::||
Db    285 EVVLSFVFWSVLTSLGLLVWYFPLWHMGISGYEACILFELSPFLLGIPLLRKFASKVPVI 344

Qy     71 TLLRIITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTVW------SGSHLQRYLRIW 124
             :|| :| || :| :  || : :: |  | :|:::|| | |      |:|| |:|:
Db    345 FL--FLNVIGIAYKLEDDVHRLFVTAF----SVCCECLAWTSLFSNISPENLAIERKIS 398

Qy    125 GFILGQIVLVLVLRIWYTSLNPIW------SYQMSNKVILTLSAIATLDRIGTDGDCSK 176
             ::| :::| |: |: | | | ||      |: ||: :: |  |||:| :||: |  |
Db    399 TFLFGLLASSIAKYSFFSNNPIWPILNETNGGKOIPALIVGIIACLIFAIFHVQOTTANA 458

Qy    177 PEERKTGEVATGMASRPNWLLAGAAFGSLVFLIFTHWVFGEVSLVSRWAVSGHPHGPDNP 236
             || |  : |: |   : :|  |:||| :||:  | ||::  || |:| | |  || | |
Db    459 VEHFKLRKITA------LSAALSLGTVLFCLHTFLCDSTVLMTWSWDGYPIKGPQPYP 510

Qy    237 EGGAVLLCLASGLMLPSCLWFRGTGLIWWVTGTASAGLLYL-HTWAAAVSGCVFAIFT- 294
             :||| |   :|: : |  ||| | :|  || |:|: ||||  |||:| :|  ||:|
Db    511 HGAVSIVSICAVLVAPYLYQSGAFMLIGFVLACFGSYFMYIINHGWCSYLGGLIFTSVVL 570

Qy    295 ----ASMWPQTLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFEVPGGVVARERSD 350
                 :::  || : || |    | :::| |: || ||:| |: || ||  |:  | |::
Db    571 IYSFPASI-----RISSFYSPAKVWGGAFLVILYSLAHVWVVAYEFVPGGPILRERTS 623

Qy    351 VLLGTMMIIGLNMLF--------GPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLL 400
             |:|| ::: || |:|        | ||: :||  |:|| :||  |::: |  | ::
Db    624 YIL----IFIGWNLAALVPAYSGGESKEPNKADSSVVDIKQSDSSYRRRSFKKS-----LL 674
```

```
Qy    401 VGVGLLGLGLR 411
             |   |: |:
Db    675 TGFCLALMALK 685
```

```
Q9S125
ID   Q9S125;         PRELIMINARY;      PRT;       279 AA.
AC   Q9S125;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative secreted protein.
GN   SCO0131 OR SCJ21.12C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RP   SEQUENCE FROM N.A.
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL109747; CAB52358.1; -.
DR   InterPro; IPR005135; Exo_endo_phos.
DR   Pfam; PF03372; Exo_endo_phos; 1.
SQ   SEQUENCE   279 AA;  30031 MW;  B16719A1698580C0 CRC64;

Query Match             3.8%;  Score 141.5;  DB 16;  Length 279;
Best Local Similarity   24.9%;  Pred. No. 0.019;
Matches     63;  Conservative   36;  Mismatches  91;  Indels  63;  Gaps  12;

Qy    447 LERSAHLLNETGADPITILESDASKPYMGNNDLTMW------LGEKLGFYTDFG----- 494
             |:|: ::  ||  || |:|:| : :||::|||  |      || :||:||| |
Db     58 LRRVANVIRKSGADVVGLQEVD-KHYSARSD--WADQPAELAELLGTHVVFGANIDNS 112

Qy    495 ---PSTRYHTWGIMALSRYPIVKSEH-HLLPSPEGEIAPAITLTVNISGKLVDFVVTHF- 549
                |:  |:||  |:| ||| ||::| |||||   | ||| || ||||||| ||||| |
Db    113 PPAPGGHRVQYGTAILSRYPITASDNTWLYKSPGQEORGLLHATLDVHGKKVEFYNTHLA 172

Qy    550 -GNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDH 608
              ||:|||||||||||||||||||||||||||||||:||||||||||||:|||||||||||
Db    173 AGSQADRLQQTAQVVDLIGTRKPG----ILVGDFNALPAAPE-------SRPLQNAYT 219
```

```
Qy   609 DRWCEYIMYRG------------LIRLGYARISHAELSDSSIQMAKFRIPDDPTNYRDNQ 656
         :||                     |||                      ||||
Db   220 DAWAKSLHARGDGATYPAQSPPTERIDLIYAT---RRVTPLVAQVLK----DDP------- 265

Qy   657 KVVIDHREVSEKI 669
         |||:|
Db   266 -AASDHRPLLGKV 277
```

RESULT 8
Q824F6
```
ID   Q824F6       PRELIMINARY;   PRT;   466 AA.
AC   Q824F6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN   TY2913.
DE   GabA permease (4-amino butyrate transport carrier).
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AL627276; CAD05902.1; -.
DR   InterPro; IPR002293; AA/rel_prmease1.
DR   InterPro; IPR004840; AaC_permease.
DR   InterPro; IPR004841; Permease.
DR   Pfam; PF00324; aa_permeases; 1.
DR   PROSITE; PS00218; AMINO_ACID_PERMEASE_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   466 AA;   50852 MW;   A7CDA258CE3D42D2 CRC64;
```

Query Match
Best Local Similarity   20.4%;   Score 133;   DB 16;   Length 466;
Matches 102;   Conservative 76;   Mismatches 151;   Indels 172;   Gaps 25;

3.68;

```
Qy     6 REILLESLLGCVSWSLYHDLGPMIYFPLQTLELTGLEGFSIAFLSPIFLTTPFWKLVN 65
         |   |||  || |  ||  | ||||
Db    17 RHVTMLSIAGVGASLF--VGSSV                                      57

Qy    66 KKWMLTLLRITIIGSIASFQAPNAKLRMVLALGVSSSLIVQAVTWWSGSHLQRYLRIWG 125
         ::|    ::| |  : ::                    ||::
Db    58 ---VVMIMRMLAEMAVAT--PDT---------GSFSTYADKAIGPWAG--YTIGWLYWWF 101

Qy   126 FIL-----GQIVLVLRIWYTSLNPIWSYQMSNKVILTLSAIATLDRIGTDGDCSKPEEK 180
         ||       :                              | ::|
Db   102 WVLVIPLEANIAAILLNSWIPGI-PVWLFSLVITLALTGSNLLSV--------------K 146

Qy   181 KTGEVATGMA-SRPNWLLAGAAFGSLVFLTHWFGEVSLVSR-WAVSGHPHPGPDPNPFG 238
         | ||      |||||||||||||||||||||||||||    ||||||||||||||||||
Db   147 NYGEFEFWLALSKVIATLAFIALGAAAISGFYPYAEVSGISRLW-----DHGGFMPNGFG 201

Qy   239 -----GAVLLCLASG---------LMLPSCLW*---------FRGTGLIW 264
         |    |||||||||          |||||||          ||||||||
Db   202 AVLSAMLITMFSFMGAEIVTIAAAESDTPDKHIVRATNSVIWRISIFYLCSIFVVALIP 261

Qy   265 W------------VTGTASAAGLLYLHTWAAAVSGCVF---------AIFTASMWPQTLGH--- 304
         |            ::|                            |  ::||
Db   262 WNMPGLKSVGSVGSYSRVSVLELLHI-PHAKFIMDCVILLSVYTSCLNSALYTASRMLYSLSRRGD 320
```

```
Qy   305 --LINSGTNPGKTMTIAMI--------FYLLEIFPCAWCTAFKFVPGGVYARERSDVLLGTM 356
           |:|   ||  | ||          ||| || |||  ||| ||| |
Db   321 APAIMGKTNRSKTPWPWAVLLSTGAAFLTVINNYAPAKVFKFL------IDSSGAIALLV 374

Qy   357 MLIIGLNMLFGPKNLDLLLQTKNSSKVLFRK-SEKYMKLFL-----WLLVG-------- 402
         || | ||||||||   ||  |  ||| ||      |||||      |||||
Db   375 YLVIAISQL----------------RMRKILLAQGGEIKILKMWLXPWLIWLVIGFICFVLVV 420

Qy   403 ---------------------VGLLGIGL 410
                              ||||||||
Db   421 MLFRPAQQLEVISTGLLGLGI          441
```

RESULT 9
Q950T2
```
ID   Q950T2       PRELIMINARY;   PRT;   583 AA.
AC   Q950T2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   NADH dehydrogenase subunit 5 (EC 1.6.5.3).
OS   Hyaloraphidium curvatum.
OG   Mitochondrion.
OC   Eukaryota; Fungi; Fungi incertae sedis; Hyaloraphidium.
OX   NCBI_TaxID=82268;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Forget L., Ustinova J., Wang Z., Huss V.A.R., Lang F.B.F.;
RT   "Hyaloraphidium curvatum: a linear mitochondrial genome, tRNA editing,
RT   and an evolutionary link to lower fungi.";
RL   Mol. Biol. Evol. 0:0-0(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Lang F.B.F.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR   EMBL; AF402142; AAK83434.1; -.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
KW   Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ   SEQUENCE   583 AA;   64424 MW;   2B568D4847987022 CRC64;
```

Query Match
Best Local Similarity   19.4%;   Score 132;   DB 8;   Length 583;
Matches 116;   Conservative 91;   Mismatches 194;   Indels 196;   Gaps 28;

3.5%;

```
Qy     8 ILLLESLLGCVSWSLYHDLGPMIYF--PLQTLELTGLEGFSIAFLSPIFLTTPFWKLVN 65
         | | |  ||  | ||  | |||     || | |   ||  |   | |
Db     5 IIILPLLGSLSSGL--LGRWLGYSGGPLLITVLLMGV--                       52

Qy    66 KKWMLTL---LRIITIGSI--ASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGS-HLQ 118
         :                  ||| || |||| ||| ||  |  ::| | :
Db    53 HQSSLLLPLPISWLTLGWLDLDMSFIMDELSISMMIPICTVSFLVHMYAIGYMIGDPHIQ 112

Qy   119 R---YLRIWGFIL--------------GQIVLVLRIWYTSLNPIWSYQMSNKVI 156
         |   |   ||                |||||||||||||||||||||||||||
Db   113 RFFSYLSLFTFFMLLMVTADNWLLLFIGWEGVGLVSYLLIGFWPTRLR--ACQAAALQAF 169

Qy   157 LTLSAIATLDRIGTDGDCSKPEKKTGEVATGMASRPNWLLAGAAFGSLVFLTHWFGEV 216
         |                            ||       | |
Db   170 L-------MNRIGDTG----------------------------LFLAMA-------- 193

Qy   217 SLVSRWAVSGHPHPGPDPNPFGGAVLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLL 276
         ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   194 EFSTFFALLPHMNP-----------------YWTALIGLLMLLAVPAK-SGQL 228

Qy   277 YLHTW-AAAVSGCVFAITTASMWPQTLGHLINSGTNPGKTMTIAMIFYLLE----IF--- 328
         || |  ||||||||| ||||||||||||                      |   ||
Db   229 GLHAWLPVAMEG----------------TMVTAGIYLLLRFNALFGFSP 272
```

Qy 329 FCAWCTAFKFVPGGVYARERSDV------------LLGTMMLIGLNMLFGPKKNDLLLQT 378
        ||| |:| | | : |||| ||||||
Db 273 ILIWLGAATALFGAVYGVGVQTDLKRTIAYSTTSQLGYMVLACGLG------------ 317

Qy 379 KNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYE-----RKIGKVAPTKEV----- 428
              : | ||:||| | |||||
Db 318 QYGLALAHLVNHAFFKALLFLSAG-----SVIHAIHDEQDIRKMGGIMSMMPVTYITSM 371

Qy 429 ----SAAIWPFRFGYDNEGW----SSLERSAHLLNETGADFITILES------------DAS 470
            |: ||| || || |::|: |||:
Db 372 VGSLSLIVALPFLTGYYSKDFILQTAFGTTGYLMG-LGAAYFTAFYSLKLLHRVFWLPPQS 430

Qy 471 KPYMGNNDLTMWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEI 527
         | | : | | | : ||: | :| : ||| |:|
Db 431 KVVLGAHEPSAWMLIPTTLTLFSIS-----NGYLAQNHAALAPAMNTLFLLPPQNQI 482

RESULT 10
Q8SPK7

ID   Q8SPK7   PRELIMINARY;   PRT;   659 AA.
AC   Q8SPK7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Sodium iodide symporter.
GN   NIS.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=THYROID;
RA   Selmi-Ruby S., Rousset B.;
RT   "Molecular cloning and functional analyses of pig sodium iodide
RT   symporter: evidence for three forms generated by alternative
RT   splicing.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ277989; CAC81949.1; -.
SQ   SEQUENCE   659 AA; 70125 MW; 3C06C869I435EA68 CRC64;

Query Match   3.5%; Score 129; DB 6; Length 659;
Best Local Similarity   21.9%; Pred. No. 0.48;
Matches   68; Conservative   42; Mismatches 107; Indels   94; Gaps   14;

Qy   15 GCVSWSLYHDLGPMI-------YYFPLQTL---ELTGLEGFSIAFLSPIFITITPFWK 62
        | : || | | | || | ||: |:| :|
Db 300 GIVMFALYVDCDPLLAGHISAPDQYMPLLVLDIFEDLPGVPG-------LFLACAYSGT 351

Qy   63 LVNKKWMLTLLRIITIGSIASFQAPN`AKLRLMVLALGVSSSLIVQAVTWWSGSHLQRYL 121
        | | : : | :|||| | :: |:|:   :|||| || | :
Db 352 LSTASTSINAMAAVTVEDLIKPRLPNLAPRRLVIISKGLSCLTVAALSSLLGGVLQCSF 411

Qy 122 RIWGFI----------LGQIVL-----VVLRIWY---TSLNPIWSYQMSNKV 155
          | | || || | | :|| | :| ::
Db 412 TVMGVISGPLLGAFVLGMFLPSCNTSGVLSGLAAGLALSLWVAVGASLYPPSAQSMG--- 468

Qy 156 ILTLSAIATLDRIGTDGDGDCSKPEEKKTG---------EVATGMASRPNWLLAGAAFGS 204
        : ||:|| || | | | || | : |:|
Db 469 VLPSSA--------AGCALPTANASGLQDPVLAVANASSTASSLETDPEOPILAASFYA 518

Qy 205 LVFLTHHVFGEVSLVSRWAVSGHPHPGPDPNPFGGAVLLCLASGLMLPSCLWFRGTGLLIW 264
        : :| | |: | | | |||:| | : ||:|:|
Db 519 ISYLYGALGTLSTI---------------LCGALISCL-TGPTKRSAL---GPGLLW 557

Qy 265 W--VTGTASAA 273
        | | |||:|
Db 558 WDLTRQTASVA 568

RESULT 11
P95556

ID   P95556   PRELIMINARY;   PRT;   396 AA.
AC   P95556;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF396 protein.
GN   ORF396.
OS   Pseudomonas stutzeri (Pseudomonas perfectomarina).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas;
OC   Pseudomonas.
OX   NCBI_TaxID=316;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZOBELL ATCC 14405;
RX   MEDLINE=97107629; PubMed=8950369;
RA   Glockner A.B., Zumft W.G.;
RT   "Sequence and analysis of an internal 9.7-kb segment from the 30-kb
RT   denitrification gene cluster of Pseudomonas stutzeri.";
RL   Biochim. Biophys. Acta 1277:6-12(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZOBELL ATCC 14405;
RX   MEDLINE=96028114; PubMed=7588711;
RA   Palmedo G., Seither P., Korner H., Matthews J.C., Burkhalter R.S.,
RA   Timkovich R., Zumft W.G.;
RT   "Resolution of the nirD locus for heme d1 synthesis of cytochrome cd1
RT   (respiratory nitrite reductase) from Pseudomonas stutzeri.";
RL   Eur. J. Biochem. 232:737-746(1995).
DR   EMBL; Z73914; CAA98157.1; -.
SQ   SEQUENCE   396 AA; 43083 MW; EB07255ECFB1E0D3 CRC64;

Query Match   3.5%; Score 128.5; DB 2; Length 396;
Best Local Similarity   21.1%; Pred. No. 0.27;
Matches   82; Conservative   55; Mismatches 147; Indels 105; Gaps   20;

Qy   55 LTITPFWKLVNKKWML-TLLRIITI-----------GSIASFQAPNAKL------RLMVLAL 98
        |: |: |:| ::| ||| | || |: ||| | |:: ||
Db   10 LSTAPIWRLAFRPFFLAGSLYALLAIPLMNVAAWTGLWPGFQPTGGWLAWHRHEMLFGFAM 69

Qy   99 GVSSSLIVQAVTWW--------SGSHLQRYLRIW------GFILG-------QIVLVVLRI 138
        |: | : || | | ::: | | | || |:|| |||:
Db   70 AIVAGFLLTAVQOTWTGQYAPSGNRLVGLAAAVWLAAARLGWLFGLPAAWLAPLDLLFVALV 129

Qy 139 WYTSLNPIWSYQMSNK----VILTLSAIATLDRIGTDGDGDCSKPEEKKTGEVATGMASRPNWL 196
        | :   : ::| |:| |: : || | | |:| |
Db 130 WMMA-QMLWAVRQRKRNYPIVVVLSLMLGADVLILTGLLQGNDALQROGVLAGL------WL 183

Qy 197 LAG----AAFGSLV--FLTH-----------WVFGEVSLVSRWAVSGHPHP---GPDPNPFG 238
        :| | |:| | | |: | ::| | | | | | |
Db 184 VAALMALIGGRVIPFFTQRGLGKVDAVKPWVWLDVALLVGTGVTAILLHAFGVAMRPQPLL 243

Qy 239 GAVLLCLASGLMLPSCLWF-RG---TGLIW----WVTGTASAAGLLYLHTWAAAVS- 286
        | : |:: | | | || | | : :||| | | | ||:
Db 244 GLLFVAIGVGHLLRLMRWYDKGIWKVGLLWSLHVAMLWL--VVAAFGLALWHFGLLAQSS 301

Qy 287 ------GCVFAFLTASMWPQTLGH----------------LINSGTNPGKTMTIAMLF 322
              || : | ||| | :|| |:|||:|| | | |||:
Db 302 PSLHALSVGSMSGLILAMIARVTLGHTGRPLQLPAGIIGAFVLFNLGTAARVFLSVAWPV 361

Qy 323 YLLEIFFCAWCTAF-----KFVPGGVYAR 346
         |||: |||| || ||||||| ||
Db 362 GGLWLAAVCWTLAFALYWRYAPMLVAAR 390

RESULT 12
Q8ZMM0

ID   Q8ZMM0   PRELIMINARY;   PRT;   466 AA.
AC   Q8ZMM0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   APC family, gamma-aminobutyrate transport protein, RpoS
DE   dependent.

```
GN   GABP OR STM2793.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21534948; PubMed=11677609;
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
DR   EMBL; AE008827; AAL21678.1; -.
DR   InterPro; IPR002293; AA/rel_permease1.
DR   InterPro; IPR004840; AAC_permease.
DR   InterPro; IPR004841; Permease.
DR   Pfam; PF00324; aa_permeases; 1.
DR   PROSITE; PS00218; AMINO_ACID_PERMEASE_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   466 AA;  50898 MW;  1892BCE3C2404308 CRC64;

Query Match             3.4%;  Score 128;  DB 16;  Length 466;
Best Local Similarity   20.2%;  Pred. No. 0.36;
Matches   101;  Conservative   76;  Mismatches 152;  Indels 172;  Gaps 25;

Qy    6 REILLLESLLGCVSWSLYHDLGPMIYYPPLQTLELTLGLEGFSIAFLSPIFLTITPPFWKLVN 65
Db   17 RHVTMLSIAGVIGASLF--VGSSV--------------AIAEAGPAVLLAYLFAGLL- 57

Qy   66 KKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGSHLQRYLRIWG 125
Db   58 --VVMIRMLAEMAVAI---PDT--------GSFSTYADKAIGPWAG-YTIGWLYWWF 101

Qy  126 FII------GQIVLVLRIWYTSLNPIWSYQMSNKVILTLSAIATLDRIGTDGDCSKPEEK 180
Db  102 WVLVIPLEANIRAIILNSWIPGI-PVWLFSLVITLALTGSNLLSV-----------K 146

Qy  181 KTGEVATGMA-SRPNWLLLAGAFGSLVFLTHWVFGEVSLVSR-WAVSGHPHPGPDNPEFG 238
Db  147 NYGEFEFWLALCKVIAILAFIALGATAISGFYPYAEVSGISRLW----DHGGFMPNGFG 201

Qy  239 --------------LMLPSCLW---------FRGTGLIW 264
Db  202 AVLSAMLITMFSFMGAEIVTIAAAESDTPDKHIVRATNSVIWRISIFYLCSIFVVALIP 261

Qy  265 W------VTGTASAAGLLYLTHWAAAVSGCVF-------AIFTASMWPQTLGH-- 304
Db  262 WNMPGLKSVGSTRSVLELLHI-PHAKFIMDCVILLSVTSCLNSALYTASRMLYSLSRGD 320

Qy  305 --LINSGTNPGKTMTIAMI------FYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTM 356
Db  321 APAIMGKINRSKIPWVAVLLSTGAAFLIVIVNYAPAKVFKFL-----IDSSGAIALLV 374

Qy  357 MLIIGLNMLFGPKKNLDLLLQTKNSSKVLFRK-SEKYMKLFL----WLLVG- 402
Db  375 YLVIAISQL---------RMRKILLAQGGEIKLKMWLYPWLTWLVIGFICFVLVV 420

Qy  403 --------------VGLLGLGL 410
Db  421 MLFRPAQOLEVISTGLLGLGI 441

RESULT 13
ID   Q82R15    PRELIMINARY;    PRT;   487 AA.
AC   Q82R15;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   DASS family, Citrate:succinate transport (antiport) protein.
GN   CITT OR STM0618.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
DR   EMBL; AE008724; AAL19569.1; -.
DR   InterPro; IPR001898; Na/sulf_symp.
DR   Pfam; PF00939; Na_sulph_symp; 1.
DR   TIGRFAMs; TIGR00785; dass; 1.
KW   Complete proteome.
SQ   SEQUENCE   487 AA;  52939 MW;  D1342A24ECB777F CRC64;

Query Match             3.8%;  Score 124.5;  DB 16;  Length 487;
Best Local Similarity   22.9%;  Pred. No. 0.69;
Matches   83;  Conservative   42;  Mismatches 133;  Indels 105;  Gaps 18;

Qy   52 PIFLTITPPFWKLVNK---------------KWMLTLLRIITIGSIASFQAPNAKLRLMVLALGV 100
Db  169 PVIKNLPPLFKSFPNDPSARRIGGYLMWMWVISTSLSSSMFVTGAAPN-----VIGLEF 222

Qy  101 SSSLIVQAVTWWSGSHLQRYLRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVILTLS 160
Db  223 VSKIAGVQISW-----LQWFL--SFLPVGIILLIVAPWL------SYVL------- 258

Qy  161 AIATLDRIGTDGDGDCSKPEEKKTGEVAT-------GMASRPNWLLLAGAAFGSLVFLTHW 211
Db  259 --------YKPEVTHSAEVAAWAGGELKNMGRLSRKEWTLIGLV---LLSLGLW 301

Qy  212 VFG------EVSLVSRWAVS----GHPHPGPDPNPFGG-----AVLLCLASGLMLPSCL 255
Db  302 VFGGKTINATAVGLLAVSLMLALHVVPWKIDTRYSNASAWNTLVNLATLVVMANGLITRSGFI 361

Qy  256 -WFRGTGLIWWVTGTASAAGLLYL-----HTWAAAVSGCVFAIFTASMWPQTLGHLIN 307
Db  362 DWFANT-MSTHLEGFSPDATIVIVLVFYFAHYLFASLSA---HTAIMLPVILA--VG 413

Qy  308 SGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGV-----YARERSDVLLGTMMLLIGL 362
Db  414 KGI-PGVPMEQLCILLVLSIGIMGCLTPYATGPGVIIYGCGYVKSRDYWRLGAIFGVIYI 472

Qy  363 NML 365
Db  473 AML 475

RESULT 14
ID   O79569    PRELIMINARY;    PRT;   609 AA.
AC   O79569;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   NADH dehydrogenase subunit 5.
GN   NADH5.
OS   Mustelus manazo.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes; Triakidae;
OC   Mustelus.
```

```
-OX  NCBI_TaxID=79736;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE=99083431; PubMed=9866199;
RA   Cao Y., Waddell P.J., Okada N., Hasegawa M.;
RT   "The complete mitochondrial DNA sequence of the shark (Mustelus
RT   manazo): Evaluating rooting contradictions to living bony
RT   vertebrates.";
RL   Mol. Biol. Evol. 15:1637-1646(1998).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DE   -1-
EMBL; AB015962; BAA33045.1; -.
DR   InterPro; IPR003916; NADHub_oxred5.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR001516; Oxidored_q1.
DR   InterPro; IPR000215; Serpin.
Pfam; PF00361; oxidored_q1; 1.
Pfam; PF00662; oxidored_q1_N; 1.
PRINTS; PR01434; NADHDHGNASES.
PROSITE; PS00284; SERPIN; UNKNOWN_1.
KW   Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ   SEQUENCE  609 AA;  68215 MW;  97FBE6D34CDE7DD0 CRC64;

Query Match             3.3%;     Score 124;    DB 8;    Length 609;
Best Local Similarity  19.3%;     Pred. No. 1;
Matches  131;  Conservative  83;  Mismatches 210;  Indels 254;  Gaps  32;

Qy  155 VILTLSAIATLDRIGTDGDCSKPEEKKTGEVATGMASRPNW--LLAGAA-----FGSLV- 206
Db   14 IILTLPLISSL---------------------SPKELK-----------PNWSSLYVKTAVKISFFISLIP 52

Qy  207 ---------FLTHW------------VFGEVSLVSRWAV----SGHHP 230
Db   53 LFIFLDQGLESIVTNWNWMNMGPPDINMSFKFDLYSIIFTPVALIYVTWSILEFALWYHHS 112

Qy  231 GPDPNPFGGAVLLCLASGLMLPSC----------LWFRGTGLI------WWVT---GTASAA 273
Db  113 IDPNNNRFKYLLLFLISMIILVTANNMFQLFIGWEGGVGIMSFLLIGWWYSRADANTAALQ 172

Qy  274 GLLY-----------LHTW-----------AAAVSGCVFAI 292
Db  173 AVIYNRIGDVGLILSMAWLATNLNSWEIHQLFILSKNKDLTLPLLGLVLAANGKSAQFGL 232

Qy  293 F-----TASWPQTLGHLINSGTNPGKTMTIAMIFYLLEIF---------FCAWCTAFK 337
Db  233 HPWLPSAMEGPTPVSALLHS-----STWVAGIFLLIRLHPLIQODNKILTVCLCLGALT 287

Qy  338 FVPGGVVARERSDV-----------LLGTMMLIIGLNMLFGPKKNIDLLLQTKNSSKYLFR 387
Db  288 TLFTATCALTQNDIKKIVAFSTSSOLGLMMVTIGLN------------QPOLAFLHI 332

Qy  388 KSEKYMKLFLWLLVGVGLLGLGLRHKAYE------RKLG---KVAPTKEVSAAIWPFRFGYD 440
Db  333 CTHAFFKAMLFLCSG-----SIIHSLNDEQDIRKMGGLHKLLPFTSTRSLTIGS----- 380

Qy  441 NEGWSSLERSAHLLNETGADFIT-ILESDASKPYMGNNDLTMWLGEKLGFYTDFGPSTRY 499
Db  381 ------------LALTGMPFLSGFFSKDAIIESMNTSHLNAWALILTLVATSF---TAI 424

Qy  500 HTWGIM--ALSRYP-----IYKSEHHLLPSP----EGEIAPAITLTVNIGKKLVDEVV 546
Db  425 YSLRLIFFALMNYPRFNTLSPINENNPLVINPIKRLAYGSIIAGLIITLNLT------- 476

Qy  547 THFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQITEHGNVKDIDST 606
Db  477 ------------PIRKTQIMTMSPLLKLSALLVTIMGLLA------LELTNLIN------ 512

Qy  607 DHDRWCEYIMYRGLIR-LGY-ARISHAELSDSSEIQMAKF---RIPDDPTNYRDNQKVVID 661
Db  513 SHFKTNPTLHYHHFSNLLGYFPSIIHRLLPKTSLNWAQYISTHLIDQTWNEKIGPKSNLI 572

Qy  662 HREVSEKIHFNPREFGSYK 679
```

```
Db   573 QQTSLIKLSTOPQQGLIK 590

RESULT 15
ID   Q91VY8
AC   Q91VY8;
         Q91VY8         PRELIMINARY;       PRT;     269 AA.
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Aquaporin 1.
GN   AQP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST TUMOR;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC007125; AAH07125.1; -.
DR   MGD; MGI:103201; Aqp1.
DR   InterPro; IPR000425; MIP_family.
DR   Pfam; PF00230; MIP; 1.
DR   ProDom; PD000295; MIP_family; 1.
DR   TIGRFAMs; TIGR00861; MIP; 1.
DR   PROSITE; PS00221; MIP; UNKNOWN_1.
KW   Porin.
SQ   SEQUENCE  269 AA;  28823 MW;  80534DE4B78AB5E7 CRC64;

Query Match             3.3%;     Score 123.5;  DB 11;  Length 269;
Best Local Similarity  22.8%;     Pred. No. 0.37;
Matches  60;  Conservative  36;  Mismatches  82;  Indels  85;  Gaps  11;

Qy   60 FWKLVNKKWM-LTLLRIITIGSIASFQAPNAKLRLMV-------LALGVSSLIVQAVTW 111
Db   10 FWRAVVAEFLAMTLFVFISTGSALGFNYPLERNQTLVQDNVKVSLAFGLSIATLAQSVGH 69

Qy  112 WSGSHLQRYLRIWGFILGQIVLVLVLRIWV--------TSLNPIWSYQMSNK----- 154
Db   70 ISGAHLNPAVTL-GLLSCQISILRAVMYIIAQCVGAIVATAILSGITISSLVDNSLGRND 128

Qy  155 --------------VILTLSAIATLDRIGTDGDCSKPEEKKTGEVATGMASRPNNL 196
Db  129 LAHGVNSGQGLGIEIIGTLQLVLCVLATTDRRRDLGGSAP-----LAIGLSVALGHL 181

Qy  197 LA--------GAAFGSLV----FLTHWVFGEVSLVSRWAVSGGHPHPGPDPNPF-GGAV 241
Db  182 LAIDYTGCSINPARSFGSAVLTRNFSNHWTF--------------WV-----------GPFIGGAL 222

Qy  242 LICLASGLMLPSCLWFRGTGLIW 264
Db  223 AVLIYDFILAPRSSDFTDRMKVW 245
```

Search completed: January 7, 2003, 22:27:46
Job time : 95 secs

124.

ALIGNMENTS

RESULT 1
AK025164
LOCUS   AK025164
DEFINITION   Homo sapiens cDNA: FLJ21511 fis, clone COL05748.   2486 bp   mRNA   linear   PRI 29-SEP-2000
ACCESSION   AK025164
VERSION   AK025164.1 GI:10437625
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL05748.
  ORGANISM   Homo sapiens

THIS PAGE BLANK (GPO)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: January 7, 2003, 13:40:54 ; Search time 5427 Seconds
                 (without alignments)
                 11277.538 Million cell updates/sec

Title: US-09-809-638-1
Perfect score: 2103
Sequence: 1 atgaactcgctggagaga.........ccaaatactttatgaaac 2103

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1455140278 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
                 1: gb_ba:*
                 2: gb_htg:*
                 3: gb_in:*
                 4: gb_om:*
                 5: gb_ov:*
                 6: gb_pat:*
                 7: gb_ph:*
                 8: gb_pl:*
                 9: gb_pr:*
                 10: gb_ro:*
                 11: gb_sts:*
                 12: gb_sy:*
                 13: gb_un:*
                 14: gb_vi:*
                 15: em_ba:*
                 16: em_fun:*
                 17: em_hum:*
                 18: em_in:*
                 19: em_mu:*
                 20: em_om:*
                 21: em_or:*
                 22: em_ov:*
                 23: em_pat:*
                 24: em_ph:*
                 25: em_pl:*
                 26: em_ro:*
                 27: em_sts:*
                 28: em_un:*
                 29: em_vi:*
                 30: em_htg_hum:*
                 31: em_htg_inv:*
                 32: em_htg_other:*
                 33: em_htg_mus:*
                 34: em_htg_pln:*
                 35: em_htg_rod:*
                 36: em_htg_mam:*
                 37: em_htg_vrt:*
                 38: em_sy:*
                 39: em_htgo_hum:*
                 40: em_htgo_mus:*
                 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	% Length	DB	ID	Description
1		2098.2	99.8	2486	9	AK025164	AK025164 Homo sapi
2		1511	71.8	2412	10	BC016523	BC016523 Mus muscu
3		267.4	12.7	184206	9	AC020593	AC020593 Homo sapi
4		192.6	9.2	218040	2	AC108828	AC108828 Mus muscu
5		191	9.1	162312	2	AC114455	AC114455 Rattus no
6		160	7.6	222885	2	AL158816	AL158816 Homo sapi
7		151	7.2	222885	2	AL158816	AL158816 Homo sapi
8		132.4	6.3	316613	8	SCCHRIII	SCCHRIII Sequence
9		122.6	5.8	218040	2	AC108828	AC108828 Mus muscu
10		121.8	5.8	1234	8	D89130	D89130 Schizosacch
11		121.8	5.8	34034	8	SPAC688	AL355632 S.pombe c
12		112	5.3	305	8	AX337478	AX337478 Sequence
13		110.6	5.3	995	11	CNS06JWV	AL402149 T3 end of
14		108	5.1	162312	2	AC114455	AC114455 Rattus no
15		81.6	3.9	170472	2	AC105576	AC105576 Rattus no
16		81.4	3.9	1348	9	HUMTPIPSA	HUMTPIPSA Human TPI (
17		59.8	2.8	1711187	7	AC116960	AC116960 Dictyoste
18		54	2.6	54	6	AX106346	AX106346 Sequence
19		54	2.6	54	6	AX140637	AX140637 Sequence
20		54	2.6	54	6	AX200497	AX200497 Sequence
21		54	2.6	54	6	AX267153	AX267153 Sequence
22	c	42.6	2.0	178874	2	AC095821	AC095821 Rattus no
23	c	42.4	2.0	68261	2	AC115084	AC115084 Homo sapi
24	c	42	2.0	162349	2	AC006226	AC006226 Homo sapi
25	c	42	2.0	190000	2	AC007072	AC007072 Homo sapi
26	c	42	2.0	292987	2	AC110450	AC110450 Homo sapi
27	c	41.6	2.0	267698	2	AC123071	AC123071 Homo sapi
28	c	41.4	2.0	161990	2	AC120482	AC120482 Rattus no
29	c	41.4	2.0	167750	2	AC016439	AC016439 Homo sapi
30	c	41.4	2.0	172710	2	AC021928	AC021928 Homo sapi
31	c	41.2	2.0	141577	9	AL606748	AL606748 Human DNA
32	c	41	1.9	6563	10	RATCYP450A	M26129 Rat cytochr
33	c	41	1.9	208805	10	AL627072	AL627072 Mouse DNA
34	c	41	1.9	286358	2	AL713872	AL713872 Mus muscu
35	c	40.8	1.9	408	11	G51484	G51484 SHGC-82855
36	c	40.8	1.9	42114	3	U50193	U50193 Caenorhabdi
37	c	40.8	1.9	110000	2	AC117400_2	Continuation (3 of
38	c	40.8	1.9	124067	9	AP000435	AP000435 Homo sapi
39	c	40.8	1.9	147358	9	AP002345	AP002345 Homo sapi
40	c	40.6	1.9	95146	2	AC099240	AC099240 Rattus no
41	c	40.6	1.9	161892	2	AC123875	AC123875 Mus muscu
42	c	40.6	1.9	195790	2	AC124505	AC124505 Mus muscu
43	c	40.4	1.9	100061	2	AC016551	AC016551 Homo sapi
44	c	40.4	1.9	170569	2	AC108109	AC108109 Homo sapi
45	c	40.4	1.9	191132	9	AC010282	AC010282 Homo sapi

ALIGNMENTS

RESULT 1
AK025164
LOCUS   AK025164   2486 bp   mRNA   linear   PRI 29-SEP-2000
DEFINITION   Homo sapiens cDNA: FLJ21511 fis, clone COL05748.
ACCESSION   AK025164
VERSION   AK025164.1 GI:10437625
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL05748.
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
               Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
               Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.

```
TITLE       NEDO human cDNA sequencing project
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 2486)
AUTHORS     Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE       Direct Submission
JOURNAL     Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1..2486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="COL05748"
                     /tissue_type="colon"
                     /clone_lib="COL"
                     /note="cloning vector pME18SFL3"
     CDS             183..2282
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15080.1"
                     /db_xref="GI:10437626"
                     /translation="MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEG
                     FSIAFLSPIFLTITPFWKLVNKKWMLTLRIIITIGSIASFQAPNAKLRLMVLALGVSS
                     SLIVQAVTWMSGSHLQRYLRIWGFILGQIVLVLARIWYTSLNPIWSYQMSNKVILTLS
                     AIATLDRIGTDGDCSKPEEKTWGEVATGMASRPNWLLAGAAFGSLVFLTHHVFGEVSL
                     VSRWAVSGHPFGPDPNPFGGAVLCLASGLMLPSCLWFRGTGLIWWVTGPASAAGLL
                     YLHTWAAAVSGCVFAIFTASMWPQTLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCT
                     AFKEVPGGVVARERSDVLLGTMMLIIGLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKY
                     MKLFLWLLVGVGLLGLGLRHKAYERKLGKVAPTKEVSAAINPFEFGYDNEGWSSLERS
                     AHLLNETGADEPITIILESDASKPYMGNNDLTMWLGEKLGFYTDFGPSTRYHWGIMALS
                     RYPIVKSEHHLLPSPEGEIAPAIILTVNISGKLVDDVVTHFHGNHEDDLDRKLQAIAVS
                     KLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDHDRWCEYIMYRGLIRLG
                     YARISHAELSDSEIQMAKFRIPDDPYNYRDNQKVVIDHREVSEKIHFNPRFGSYKEGH
                     NYENNHINFHMNTPKYFL"
BASE COUNT      692 a    516 c    606 g    672 t
ORIGIN
Query Match          99.88;    Score 2098.2;    DB 9;    Length 2486;
Best Local Similarity 99.9%;   Pred. No. 0;     Mismatches 3;  Indels 0;  Gaps 0;
Matches 2100; Conservative .0;  Mismatches .0; Indels 0;

Qy         1 ATGACCTCGCGTGTGGGAGAGAAAATCCTCTTGGAGTCGCCTGCTGGGAGATGTGTTTCTTGGTCT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       183 ATGACCTCGCGTGTGGGAGAGAAAATCCTCTTGGAGTCGCCTGCTGGGAGATGTGTTTCTTGGTCT 242

Qy        61 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACAACTAGAAACTCACT 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       243 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACAACTAGAAACTCACT 302

Qy       121 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTTCTTCTCCAATATTCCTAACAATTACTCCTTTC 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       303 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTTCTTCTCCAATATTCCTAACAATTACTCCTTTC 362

Qy       181 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       363 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 422

Qy       241 ATAGCCTCCTTCCCAGGCTCCAAATGCCAAAACTTCGACTATATTCGTCGCTTGGGGTG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       423 ATAGCCTCCTTCCCAGGCTCCAAATGCCAAAACTTCGACTATATTCGTCGCTTGGGGGTG 482

Qy       301 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTCGGTGGGAAGTCATTTGCAAAGGTAC 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
...
```

```
Db       483 TCTTCCTCACTGATAGTGCCAAGCTGTGACTTGGTCGGTGGGAAGTCATTTGCAAAGGTAC 542

Qy       361 CTCAGAATTTGGGGATTCATTTTAGGACAGAGATTGTTCTTGTTGTTGTTCTACGCATAGGTAT 420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       543 CTCAGAATTTGGGGATTCATTTTAGGACAGAGATTGTTCTTGTTGTTGTTCTACGCATATGGTAT 602

Qy       421 ACTTCACTAAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       603 ACTTCACTAAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 662

Qy       481 GCCATAGCCACACTTGATCGTATTGGACAGATGGTGCAGTAAACCTGAAGAAAG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       663 GCCATAGCCACACTTGATCGTATTGGACAGATGGTGCAGTAAACCTGAAGAAAG 722

Qy       541 AAGACTGGTGAGGTAGCCACGGGGATAGGGCCTCTAGACCCAACTGGGCTGCCTGGCAGGGCT 600
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       723 AAGACTGGTGAGGTAGCCACGGGGATAGGGCCTCTAGACCCAACTGGGCTGCCTGGCAGGGCT 782

Qy       601 GCTTTTGGTAGGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTTCC 660
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       783 GCTTTTGGTAGGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTTCC 842

Qy       661 AGAATGGGCAGTGAGTGGGCATCCAACATCCAGGGCCAGATGGGCAGTAAACCCATTTGGAGGTGCA 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       843 AGAATGGGCAGTGAGTGGGCATCCAACATCCAGGGCCAGATGGGCAGTAAACCCATTTGGAGGTGCA 902

Qy       721 GTACTGCTGCTGCTTGGGCAAGTGGGAATGGAGTTGATGCTTCCATCTGTGGTTGTTGTGGTTACT 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       903 GTACTGCTGCTGCTTGGGCAAGTGGGAATGGAGTTGATGCTTCCATCTGTGGTTGTTGTGGTTACT 962

Qy       781 GGTTTGATCTGGTGGGTTTACAGGAACAACCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       963 GGTTTGATCTGGTGGGTTTACAGGAACAACCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 1022

Qy       841 TGGGCAGCTGCTGTCTGTCTGGCTGTGTCTTCGGCCATCCATCGTTGGTGCGCCCCAA 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1023 TGGGCAGCTGCTGTCTGTCTGGCTGTGTCTTCGGCCATCCATCGTTGGTGCGCCCCAA 1082

Qy       901 ACACTTGGACACCTTATTAACTCAGGGACGAACCCTGGGAAAACCATGACCATTGCCATG 960
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1083 ACACTTGGACACCTTATTAACTCAGGGACGAACCCTGGGAAAACCATGACCATTGCCATG 1142

Qy       961 ATATTTTATCTTCTAGAAAATATTTTTCTGTGCCTGGTGCCACAGCTTTTAAGTTTGTCCCA 1020
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1143 ATATTTTATCTTCTAGAAAATATTTTTCTGTGCCTGGTGCCACAGCTTTTAAGTTTGTCCCA 1202

Qy      1021 GGAGGTGTCTACGCTAGAGAAGAGATCAGATGGTGCTTTGGGACAAATGATTGTTAATATATC 1080
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1203 GGAGGTGTCTACGCTAGAGAAGAGATCAGATGGTGCTTTGGGACAAATGATTGTTAATATATC 1262

Qy      1081 GGGCTGAATATGCTATTTGGTCCTGAAGAAAAACCTTGACTTGCCTTCTTCATCAAACAAAAAAC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1263 GGGCTGAATATGCTATTTGGTCCTGAAGAAAAACCTTGATTTGCCTTCTTCAAACAAAAAAC 1322

Qy      1141 AGTTCTAAAGTGCTTTCAGAGAGTGAAAAAATACATGAAACTTTTTCTGTGGGCTGCTT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1323 AGTTCTAAAGTGCTTTCAGAGAGTGAAAAAATACATGAAACTTTTTCTGTGGGCTGCTT 1382

Qy      1201 GTTGGTGTGTGGGATTGTTGGGGATTAGGCGCATAAAAGCCTATGAGGAAAAACTGGC 1260
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1383 GTTGGTGTGTGGGATTGTTGGGGATTAGGACTACGGCGCATAAAAGCCTATGAGGAAAAACTGGC 1442

Qy      1261 AAAGTGGCCACCAAAGAAGTGCTTCTGCTGGCCATCTGGCCTTCAGGTTTGGATATGAC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1443 AAAGTGGCCACCAAAGAAGTGCTCTCTGCTGGCCATCTGGCCTTCAGGTTTGGATATGAC 1502

Qy      1321 AATGAAAGGTGGGTCTAGTCTAGAAAAGAGATCAGTCTCACCTGGCTCAATGAAAACAGGTGACAGAT 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1503 AATGAAAGGTGGGTTCTAGTCTAGAAAAGAGATCAGTCACCTGGCTCAATGAAAACAGGTGACAGAT 1562

Qy      1381 TTCATAACAAATTTGGAGTGATGCTTCTAGCCCTATATGGGGACAATGACTTAACC 1440
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1563 TTCATAACAAATTTGGAGTGATGCTTCTAGCCCTATATGGGGACAATGACTTAACC 1622
...
```

```
Qy   1441 ATGTGGCTAGGGAAAAGTTGGGTTTCTTATACAGAGACTTGGTCCAAGCACACAAGGTATCAC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1623 ATGTGGCTAGGGAAAAGTTGGGTTTCTTATACAGAGACTTGGTCCAAGCACACAAGGTATCAC 1682

Qy   1501 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1683 ACTTGGGGGATTATGGCTTTGTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCT 1742

Qy   1561 CCGTCACCAGGAGCCGAGATGCCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1743 CCGTCACCAGGAGGGCGAGATCCCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAG 1802

Qy   1621 CTGGTGGATTTTGTCGTGGACACTTTGGGACACCGCAAGAGATGACCCTGCCACGAGAAACTG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1803 CTGGTGGATTTTGTCGTGGACACTTTGGGACACCACCAAGAGAGATGACCTGCCACGAGAAACTG 1862

Qy   1681 CAGGCTATTGCTGTGTTTCAAAACTACTACTGAAAAGTAGCTCTAATCAAGTGATTATTTCTGGGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1863 CAGGCTATTGCTGTGTTTCAAAACTACTGAAAAACTGAGATATTTCTGGGA 1922

Qy   1741 TATATCACTTCAGCACCTGGGTCTCCAGAGATTATCTCACTGAAACATGGCAATGTG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1923 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTCACTGAAACATGGCAATGTG 1982

Qy   1801 AAGGATATCGACAGCACTGACTCTGATCATGACAGAGAGGTGTGAAATACAATTAGTATGTCGGAGGGCTG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1983 AAGGATATCGACAGCACTGACTCATGACAGAGATGGTGTGAAATACAATTATGTATGTCGGAGGGCTG 2042

Qy   1861 ATCAGGTGGGTTATGCGAAGAAATCTCCCATGCTGTGAACTGACTGAGTGAAAATTCAGATG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2043 ATCAGGTGGGTTATGCCAAGAAATCTCCCATGCTGAACTGACTGAGTGAAAATTCAGATG 2102

Qy   1921 GCAAAAATTAGGATCCCTGATGAGCCCCCACTAATTATATAGAGACAACCAGAGAAAAGTGGTCATA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2103 GCAAAAATTAGGATCCCTGATGAGCCCCACTAATTATATAGAGACAACCAGAAAGTGGTCATA 2162

Qy   1981 GACCACAGAGAAGAGTTTCTGAGAAAATTCATTTTAATCCCCAGATTTGGATCCTTACAAGAA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2163 GACCACAGAGAAGAGTTTCTGAGAAAATTCATTTTAATCCCCAGATTTGGATCCTTACAAGAA 2222

Qy   2041 GGACACAATTATGAAAAACAACCATCATTTCAATGAATACTCCCAAATAACTTTTTTATGA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2223 GGACACAATTATGAAAAACAACCATAATTTCATATGAATACTCCCAAATACTTTTTATGA 2282

Qy   2101 AAC 2103
          |||
Db   2283 AAC 2285
```

Query Match   71.8%; Score 1511; DB 10; Length 2412;
Best Local Similarity   82.4%;
Matches 1733; Conservative   0; Mismatches 370; Indels   0; Gaps   0;

```
Qy     1 ATGACCTCGCGTGTGGAGAGAAAATCCTCTTCTGGAGTCGCTCGCTGGGGATGTGTTCTTGTTGTCT 60
         || || |||||||||| ||||||||||| |||  | |||||  || | |||||||||||||||||||
Db   113 ATGCCAGGGCCTGTGGGAGGAGCCATGGCCCTGGAGACCTTGCTGGGCTACGTCTCCTGGTCT 172

Qy    61 CTCTACCATGACCTGGGGACCGGATGATCTATTACTTCCGTTTGCAAACACTCACT 120
         ||||||||| |  || |||||||| ||  ||||||   | ||||| || ||||||||||
Db   173 CTCTACCACGGACTGAGCCCCATGATCTATTATTTCCCCCTGCAAACGCTGGGCTCACA 232

Qy   121 GGGCTTGAAGGTTTTAGTATAGCATTCTTCTTCTGCCAAATATTCCTAACAATTACTCCTTTC 180
         |||  ||||| ||||||| || |||   || |||||| ||||||||||||||| ||||||||
Db   233 GGGCTCGGAATTCTTTTGGTGCTTGTTCTCCGTGTCCGGATACTCTTAACAATTACCCCACTC 292

Qy   181 TGGAAATTGGGTTAACAAGAGAGTGGATGTGCTAACCCTGCTGAGGAGATAATCACTATTGGCAGC 240
         |||||| |||  || |||||| | ||||||| || ||| | |||||| ||||| ||||||||||
Db   293 TGGAAACTCGTCAATAAGAAGTGGACACTCTCTCTGCTGAGAATCGTCACTGTAGGCAGC 352

Qy   241 ATACCCTCCTTCCAGGCTCCAAAATGCCAAAACCTTGCACTCATGGTTCTTCCGCGTTTGGGGGTG 300
         ||||| || |||||||||||||||| || || ||||||| ||||| ||| ||| |||||||||||
Db   353 ATAGCTTCCTTTGAGGCTCCAAATGCCAAGCTTCGAATGGTTCTGGCCCTTGGTGTG 412
```

```
Qy    301 TCTTCCTTCACTGATTAGTCCAAGCTTGTGACTTGGTGGTCGGGGAGTCATTTGCAAAGGTAC 360
          |||||||| || ||| ||||||  ||||||| |||  | |||||||||||||||||||||||||
Db    413 TCTTCTTCCTTGATTAGTGCAAACTGTGACCTGGTGTTGGTGGGAAGTGGATTGCAAAGGTAC 472

Qy    361 CTCAGAATTTGGGGGATTCATTTTTAGGACAGATTGTTTCTTGTTTCTTACGCATATGGTAT 420
          |||| |||||  ||||| ||| || ||| |  |||| |||||| |||| | |||||||||||
Db    473 CTCAAAAATTCGGGGATTCATTCTGGGACACGTCCTTCTCTTGTGTCCTCCGAATATGGTAC 532

Qy    421 ACTTCACTAAACCCAAACTCTGGAGTTATCAGAAGTGATAGTGACATTAAGT 480
          || | |  ||||||||||||||  || | ||||||   | |||||||||||||
Db    533 ACATCATTAAAACCCAAATTGGAGCTATCATGAGATGTTCGACATTGAGT 592

Qy    481 GCCATAGCCACACTTGATGCGTATTGGCACAGAGAGGTGACTGCAGTGTAAACCTGAAGAAAAG 540
          ||| |||||  ||||||   | ||||||||||||||||||||||||||||||||| | |||
Db    593 GCCGTACCTGTCCTTGACCGAATTGGCACAGAGGTGACTACAGAAGACCCTGACATTGAC 652

Qy    541 AAGACTGGTGAGGTAGCCACACGGGGATGGCCTCTAGACCACTGGGCTGCTGGCAGGGGGCT 600
          |||||  || || || |  ||||| |||||||||||||  |||||||||| |||||||||||
Db    653 AAGCCCCGTGAGGTTGCCCACGGGGCCAGGACCTCTTCTATCTAGTTGGCTCGTTGCCAGGCCCC 712

Qy    601 GCTTTTTGGTTAGCCTTTGTGTTCCTCCACCCCACTCGGTGTTTTGTGAGGAGTCTCTCTGTTCC 660
          | || | ||| | |||| ||||| | || ||||||| ||||||||  | |||||||||||||||||
Db    713 GCCTTTGGGAGCCTCCTTGTTCCTAACTCACTCACTCGGATGAGGTGTCCCATTGTTTTTC 772

Qy    661 AGATGGGGCAGTGAGTGGGGCATCCACACATCCAGGGAGCCTAACCCAGTTGGAGGTGTCA 720
          |||||| || || |  |||||||||  || ||| ||||||| ||||||||  || |||||||
Db    773 AGATGGGGCCAGTGCAGTGTCATCCCCACCCCAGGAATCCTAATCCTTTTTGGAGGTGCG 832

Qy    721 GTACTGCTGCTGCTTGGGCAAGTGGATTGGATGGCTTCCATCTGTTGTTGGTTTCCTGGTGACT 780
          || |||| ||| ||  |||||||||||||||||| ||||| ||||||||  || ||  |||||||
Db    833 GTTCTACTGGGCTTCTCAAGTGGACTGGTTCATCAGGCTTCATGATGCT 892

Qy    781 GTTTTGATCTGTGTGGGGGTTACAGGAACAGCTTCAGCTGGTCGCGGGGTCCTCCTTTTTACTTGCACACA 840
          |  ||| |||| |||||||||| |  ||| || ||||| |||| | ||||||||| |||| |  ||||||||
Db    893 GGGGTTAGCCTGGTGGATGACAGGAGCAGCTTCCGCCATGGGTCTCCTCTAACCTGCGGCACA 952

Qy    841 TCGGCACCTGCTGCTGTGTCGTCGGCTCTTCGCCATCTTTTACTGCATCCCATGTGTGGCCCCA 900
          || |||||||| ||| | |||||| | |||| |||| | |||||||||||||||| |||||||||
Db    953 TCGGGCACGTGCTGTGTTCTGGAGTGTCCTGGCAGTGTCTTTACGGGGTCTTTATGTCGGCTCAA 1012

Qy    901 ACACTTGGACACCTTATTAATTCACAGGGAACAAACCTGGGAAAAACCCATGAACACCATTGCCATG 960
          ||| || ||| ||  | | | |||| | ||| ||||||||| |||||| ||| |||||||||||||||
Db   1013 GTACTTCGGCCATCTTGTGAACTCAGGGAAAAACTGGGGAGAAGCCATGGCCACTGGCATG 1072

Qy    961 ATATTTTATCTTCTAGAAAAATATTTTTCTGTGTGTGTCACCA 1020
          |  || | | |  || |||||||||||||||| |  ||||||||||
Db   1073 ATCCTTTATGTTTTACAAACATTTTCTGTGCCTGGTTGCCACAGCGTTTGAAGTTTGTTCCC 1132

Qy   1021 GGAGGGTGTCTACGGCTAGAGAAAGATCAGATGTGTGCTTTTGGGGACAAIGATGTTAATTATC 1080
          |||||||||||| || || ||||||||||||  ||||| ||||||||| | ||| | ||||||||
Db   1133 GGAGGGTGTCTACGCTAGAGAAAAGATCTGATGTGCTTTTTGGGGACCATAATGTTAATTATT 1192

Qy   1081 GGGCTGGAATATGCTATTTTGGTCCTAAGGAAGAAAAAACCTTGACTTGCTTCTTCTTCAAACACAAAAAC 1140
          | || ||| ||  ||| || ||||| ||||||||||||||||||||| || || || |||| |||||||||
Db   1193 GGATTGAGTAGTCCTGTTTTGGGACCTAAGAAACTTTGACTTTTCTTCTTCAAACANAAAAAC 1252

Qy   1141 AGTTCTAAAGTGCTTTTTCAGAAAAGAGTGAAAAMATACAITGAAAACATTTTTCTGTCTGGCTCTT 1200
          ||  |||| ||| || |||||||||| ||||   || ||  |||| ||||| || || || |||||||
Db   1253 AGTCCTAAAACGCTTTTGAGATGCAGTGAAAAAATACATGAAAACTCATTTCTGTGGCTGTTT 1312

Qy   1201 GTTGGGTCGTGGGAATTGTTGGGGATTAGGACTAGCTGAGTGAGAAAACTGGGC 1260
          |||||||||| | ||  |||||||||||||||||| | ||| |||||||||||||
Db   1313 GTTGGGTGTGGGGTTACTGGGGATTAGGGCTACGGCACAGAAACCTACGAGAGGCCAGTTAGGC 1372

Qy   1261 AAAGTGGCCACCAACCAAAGAGAGTCTCTCGCTGCCATCTGGCCGTTGGATATGAC 1320
          || | ||||||||||||||||| |||| ||| || |||||||||||| |||||||||
Db   1373 CGGAGGGGCCACCGACAGTGGCTCTCCGGCTACATCTGGCCTTTTAGGTTTGGATATGAC 1432

Qy   1321 AATGAGAGGTGTTGCTCTACTGTCTAGAAAAINCAGTCACCTGCTCAATAAAACAGGTGCAGAT 1380
          | ||||||||| | ||| |||||||||| |  | ||| | |||| ||||| | | ||| |||||
Db   1433 AATGAAGGCTGGCCCAATCTAGAGAGTCTGCTCAACTGCTCAAGGACACAGGTGCAGAT 1492

Qy   1381 TTCATAACAAATTTGGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAAATGACTTAACC 1440
          |||||  || |||  |||  |  |  ||||||| |||||||| ||||| | | || |||||
```

RESULT 3

LOCUS   AC020593   184206 bp   DNA   linear   PRI 07-NOV-2001
DEFINITION Homo sapiens BAC clone RP11-317G22 from 4, complete sequence.
ACCESSION   AC020593
VERSION   AC020593.6 GI:14589709
KEYWORDS   HTG.
SOURCE   Homo sapiens.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 184206)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE   Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
  PUBMED   9847074
REFERENCE   2 (bases 1 to 184206)
  AUTHORS   Swearengen,S. and Cotton,M.
  TITLE   The sequence of Homo sapiens BAC clone RP11-317G22
  JOURNAL   Unpublished

```
Db   1493 TTTATCACCATTTTGGAGGAGCGATGCTTCCAAGCCCTATATAGGGAACAAACGACTTAACA 1552

Qy   1441 ATGTGGCTAGGCGAAAAGTTGGGTGTTTCTTATACAGACTTTGGCCAAGCACAAGGTATCAC 1500
          |||||| || |||| | |  | |||||||| | | ||  ||||| |||||| ||||||||||||
Db   1553 ATGTGGGCTCGGGGAGGAAGCTGGGTGTTTCTACACAGACTTTGGGCCAAGCACCAGGATCAC 1612

Qy   1501 ACTTGGGGGATTATGGCTTTGTCAAGATAACCAATTGTGAAAATCTGAGCATCACCTTCTT 1560
          | ||| || ||||| ||  |||||||| |  ||||||  |||||||| | |||||||||||||
Db   1613 ACCTGGGGGATTATGGTGGCTGTCTCGTGGTACCCGATTGTCAGATCGGAACATCACCTTCTT 1672

Qy   1561 CCCTCACCACGAGGGCGACGAATCCGACCACCACCACCATCAC 1620
          ||| |||| |||| | || |||| |||||||||||||| ||| ||
Db   1673 CCGTCGCCAGAGGGCGAGATCGCAGCAGGCCCATAAGCCAATTTCGGGAAG 1732

Qy   1621 CTGGTGGATTTTGTCGTGACAACACTTTGGGAACCACAGAAGAGACCTGGACAGGGAAACTG 1680
          ||| ||| ||| || |||| |||||||| ||||  ||||||||| || |||  ||||||||||
Db   1733 CTGGTGGATTTTGTGGTGGTGACACACATTTGGGAATCATGAAGATGACCTTTGACAGGAAGCTA 1792

Qy   1681 CAGGCTATTGCTGTTTTCAAAACTACTGAAAAGTAGCTCTTAATCAAGTGATATTTTCTGGA 1740
          |||||||| || ||||| ||||||  | |||||| || || |  || ||||  ||||| | ||
Db   1793 CAGGCTATTGCAGTTTCAAAACTGCTGAAAATTGTTCAAATCAAAGTGATATTTCTGGGT 1852

Qy   1741 TATATCACTTCACCACTCTGGCTCCAGACAGATTATCTCACTGAAACATGGCCAAAGTG 1800
          |||||||| || ||||||||||||||||| |||||| | ||||||| || ||||||  ||
Db   1853 TATATCACTTCAGAGCCTGGTTCCAGAGATTACATTACAAAACATCGGCCAAGGTG 1912

Qy   1801 AAGGATATCGACACCACTGATCATGACAAGATGGTGTGAATACATTATGTTATCGAGGGCTG 1860
          | ||||||| |||||||||| ||||  || ||||||||||  |||||||||||| ||||||||
Db   1913 AAGGATATTGACAGCTCAGATGGAGACAGATCATGACACCAGTAGCTCTTCAATCCCGGGGTCTG 1972

Qy   1861 ATCAGGTTGGGTTATGCAAGAATCTCCATGCTGAACTGAGTGAGTTCAGAAATTCAGATG 1920
          |||||||| || ||| |||||||||| ||||  ||||||||||||||||||||||||||||
Db   1973 ATCAGGTTGGGCTATGCAAGAATCTCTCATGCAGAACTGAGTGAGTTCAGAGAATTCAGATG 2032

Qy   1921 GCAAAAATTTAGGAITCCCCTGATGACCCCACTAATTATTATAGAGACAAAAGTGGTCATA 1980
          || ||||||| || | ||||||| ||  | | ||||| | | ||||||||| | ||||||||
Db   2033 GCCAAAATTCAGGATCCCCGGATGATCCCCGCAAATTACAGGAGACTGTCATA 2092

Qy   1981 GACCACAGAGAAGTTCTGAGAAAATTCATTTTTAAATCCCAGATTTGGATCCTACAAAGAA 2040
          ||||||||| ||||| ||||||||||| ||  |||| | |||||||| ||||||||||||||
Db   2093 GACCACCGAGGAGTTCCCAAGAACATTCACTTGCTCCCAGATTTGGGCTCCTACAAAGAA 2152

Qy   2041 GGAGACAATTATGAAAACAACCATCATTTTCATATGAATACTCCAAAATACTTTTTATGA 2100
          ||||||||| || ||||| || |||||||||||| ||| ||| || ||||||||| ||||
Db   2153 GGACACAATTATGAGAACAACCATCATTTTCACATGAATACACACCCAAAATACTTCGTTTGA 2212

Qy   2101 AAC 2103
          | |
Db   2213 ACC 2215
```

```
REFERENCE   3  (bases 1 to 184206)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 184206)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 184206)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Jul 3, 2001 this sequence version replaced gi:12057004.
COMMENT     ------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            ------------- Summary Statistics
            Center project name: H_NH0317G22
            -------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR: pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is AC011956. Actual start of this
            clone is at base position 1 of RP11-317G22.
            The clone sequenced to the right is AC001956. Actual end is at base
            position 184206 of RP11-317G22.

            The sequence fidelity of RP11-317G22 between bases 17725 to 17956
            can not be guaranteed due to an unresolved dinucleotide repeat. The
            sequence fidelity of RP11-317G22 between bases 86965 to 87015 can
            not be guaranteed due to an unresolved homopolymeric run.
FEATURES             Location/Qualifiers
     source          1..184206
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-317G22"
                     /clone_lib="RPCI-11"
     repeat_region   5..70
                     /rpt_family="L2"
     misc_feature    252..494
                     /note="match to EST AA464688 (NID:g2189572) zx82a02.s1"
     misc_feature    252..374
                     /note="match to EST AI077951 (NID:g3412359) oy39a10.s1"
     misc_feature    252..374
                     /note="match to EST AI088656 (NID:g3427715) qb14a11.x1"
     misc_feature    252..374
                     /note="match to EST AW005009 (NID:g5853713) wv02e08.x1"
     misc_feature    252..374
                     /note="match to EST AW006753 (NID:g5855531) wt07a02.x1"
     misc_feature    252..374
                     /note="match to EST BF055447 (NID:g10809343)"
     misc_feature    252..374
                     /note="match to EST BG165990 (NID:g12672693)"
     misc_feature    252..374
                     /note="match to EST BG169511 (NID:g12676214)"
     misc_feature    252..374
                     /note="match to EST BG214746 (NID:g13740767)"
     misc_feature    252..374
                     /note="match to EST BG291592 (NID:g13049843)"
     misc_feature    256..374
                     /note="match to EST W80356 (NID:g1391393) zh49a05.s1"
     misc_feature    280..374
                     /note="Similar to Bos taurus EST AV609860 (NID:g9745530)"
     misc_feature    347..374
                     /note="match to EST AW371794 (NID:g6876448)"
     misc_feature    708..1203
                     /note="match to EST AA303948 (NID:g1956281)"
     misc_feature    1021..1308
                     /note="match to EST BI063872 (NID:g14471399)"
     misc_feature    1119..1143
                     /note="match to EST T06930 (NID:g318079)"
     misc_feature    1119..1143
                     /note="match to EST AI220135 (NID:g3802338)"
     misc_feature    1119..1143
                     /note="match to EST AI271747 (NID:g3890914)"
     misc_feature    1119..1143
                     /note="match to EST BG167015 (NID:g12673728)"
     misc_feature    1119..1143
                     /note="match to EST W68728 (NID:g1377608) zd37c01.s1"
     misc_feature    1183..1538
                     /note="match to EST BG546183 (NID:g13544848)"
     repeat_region   1319..1350
                     /rpt_family="AT_rich"
     repeat_region   1349..1466
                     /rpt_family="Alu"
     misc_feature    1490..1538
                     /note="match to EST AA847380 (NID:g2938898) of07d05.s1"
     misc_feature    1490..1538
                     /note="match to EST AI220135 (NID:g3802338) qi44g05.x1"
     misc_feature    1490..1538
                     /note="match to EST AI271747 (NID:g3890914) qj77b10.x1"
     misc_feature    1490..1538
                     /note="match to EST BG167015 (NID:g12673728)"
     misc_feature    1490..1538
                     /note="match to EST W68728 (NID:g1377608) zd37c01.s1"
     misc_feature    1491..1546
                     /note="match to EST BF055447 (NID:g10809343)"
     misc_feature    1491..1538
                     /note="match to EST AA303948 (NID:g1956281)"
     misc_feature    1491..1538
                     /note="match to EST AI077951 (NID:g3412359) oy39a10.s1"
     misc_feature    1491..1538
                     /note="match to EST AI088656 (NID:g3427715) qb14a11.x1"
     misc_feature    1491..1538
                     /note="match to EST AW005009 (NID:g5853713) wv02e08.x1"
     misc_feature    1491..1538
                     /note="match to EST AW006753 (NID:g5855531) wt07a02.x1"
     misc_feature    1491..1538
                     /note="match to EST AW371794 (NID:g6876448)"
```

RESULT 4
```
LOCUS       AC108828        218040 bp    DNA     linear   HTG 28-APR-2002
DEFINITION  Mus musculus clone RP23-153H17, WORKING DRAFT SEQUENCE, 21 ordered
            pieces.
ACCESSION   AC108828
VERSION     AC108828.2  GI:20336119
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

REFERENCE   1 (bases 1 to 218040)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Mus musculus, clone RP23-153H17
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 218040)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,R., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (31-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 218040)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,R., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (28-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Apr 28, 2002 this sequence version replaced gi:18450089.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            -------------------- Project Information
              Center project name: L20971
              Center clone name: 153_H_17
            -------------------- Summary Statistics
              Sequencing vector: Plasmid; n/a; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 212760 bases at least Q40
```

Feature table:
```
     misc_feature   1491..1538
                     /note="match to EST BG165990 (NID:g12672693)"
     misc_feature   1491..1538
                     /note="match to EST BG169511 (NID:g12676214)"
     misc_feature   1491..1538
                     /note="match to EST BG214746 (NID:g13740767)"
     misc_feature   1491..1538
                     /note="match to EST W80356 (NID:g13913393) zh49a05.s1"
     misc_feature   1491..1538
                     /note="similar to EST AV609860 (NID:g9745530)"
     misc_feature   1491..1508
                     /note="similar to Bos taurus EST BG207366
                     (NID:g13729053)"
     misc_feature   1491..1508
                     /note="match to EST BG291592 (NID:g13049843)"
     misc_feature   1606..1606
                     /note="similar to Homo sapiens EST BG207366
     repeat_region   1561..1606
                     /rpt_family="AT_rich"
     repeat_region   1696..1986
                     /rpt_family="Alu"
     repeat_region   2175..2217
                     /rpt_family="Alu"
     repeat_region   2311..2780
                     /rpt_family="L1"
     repeat_region   2828..3176
                     /rpt_family="MaLR"
     repeat_region   3217..3544
                     /rpt_family="MaLR"
     repeat_region   3617..3648
                     /rpt_family="Alu"
     repeat_region   3793..4078
                     /rpt_family="AT_rich"
     repeat_region   4138..4267
                     /rpt_family="Alu"
     repeat_region   4462..4655
                     /rpt_family="Alu"
     repeat_region   4639..4688
                     /rpt_family="MER1_type"
```

```
Query Match         12.7%;  Score 267.4;  DB 9;  Length 184206;
Best Local Similarity   94.5%;  Pred. No. 1.8e-63;  Mismatches 16;  Indels 0;  Gaps 0;
Matches  277;  Conservative  0;

Qy        799 ACAGGAACAGCTTCAGCTCGCGGGCCTCCTTTCACCTGCACACAGTGGCAGCTGCTGTGTGTCT 858
              |||||||||||||||||| ||||||| ||||||| ||||||||||||||||||||||||||||
Db     111216 ACAGGAACAGCTTCAGCTGCGGGGCCTCCTTTACCTGCACACAGTGGCAGCTGCTGTGTGTCT 111275

Qy        859 GGCTGTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCCCAAACACTTGGACACCTTATT 918
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111276 GGCTGTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCCCAAACACTTGGACACCTTATT 111335

Qy        919 AACTCAGGGACAAAACCCTGGGAAAAACCATGACCATTGCCATGATATATTTTATCTTCTAGAA 978
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111336 AACTCAGGGACAAAACCCTGGGAAAAACCATGACCATTGCCATGATATATTTTATCTTCTAGAA 111395

Qy        979 ATATTTTCTGTGTCGTGGTGCACAGCTTTTTAAGTTTGTGTCCCACGAGGGTGTCTACGCTAGA 1038
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111396 ATATTTTCTGTGTCGTGGTGCACAGCTTTTTAAGTTTGTGTCCCACGAGGGTGTCTACGCTAGA 111455

Qy       1039 GAAAGATCAGATGTGCTTTTGGGGGACAATGAGTGTTAATTATATCGGGCTGAATAT 1091
              ||||||||||||||||||| ||||  |  ||  ||  ||  |  |  ||  ||
Db     111456 GAAAGATCAGATGTGCTTTTTGGGGTAGTACATTTGAAAGGTCTGGATTAAAAT 111508
```

```
Consensus quality: 214579 bases at least Q30
Consensus quality: 215279 bases at least Q20
Insert size: 220000; agarose-fp
Insert size: 216040; sum-of-contigs
Quality coverage: 7.5 in Q20 bases; agarose-fp
Quality coverage: 7.6 in Q20 bases; sum-of-contigs

------------

NOTE: This is a 'working draft' sequence. It currently
  consists of 21 contigs. Gaps between the contigs
  are represented as runs of N. The order of the pieces
  is believed to be correct as given, however the sizes
  of the gaps between them are based on estimates that have
  provided by the submittor.
  This sequence will be replaced
  by the finished sequence as soon as it is available and
  the accession number will be preserved.

*    1 7063: contig of 7063 bp in length
* 7064 7163: gap of 100 bp
* 7164 8251: contig of 1088 bp in length
* 8252 8351: gap of 100 bp
* 8352 9118: contig of 767 bp in length
* 9119 9218: gap of 100 bp
* 9219 10246: contig of 1028 bp in length
* 10247 10346: gap of 100 bp
* 10347 53500: contig of 43154 bp in length
* 53501 53600: gap of 100 bp
* 53601 55707: contig of 2107 bp in length
* 55708 55807: gap of 100 bp
* 55808 59529: contig of 3722 bp in length
* 59530 59629: gap of 100 bp
* 59630 61934: contig of 2305 bp in length
* 61935 62034: gap of 100 bp
* 62035 66831: contig of 4797 bp in length
* 66832 66931: gap of 100 bp
* 66932 72794: contig of 5863 bp in length
* 72795 72894: gap of 100 bp
* 72895 77248: contig of 4354 bp in length
* 77249 77348: gap of 100 bp
* 77349 84268: contig of 6920 bp in length
* 84269 84368: gap of 100 bp
* 84369 93794: contig of 9426 bp in length
* 93795 93894: gap of 100 bp
* 93895 107682: contig of 13788 bp in length
* 107683 107782: gap of 100 bp
* 107783 125724: contig of 17942 bp in length
* 125725 125824: gap of 100 bp
* 125825 137181: contig of 11357 bp in length
* 137182 137281: gap of 100 bp
* 137282 154808: contig of 17527 bp in length
* 154809 154908: gap of 100 bp
* 154909 181339: contig of 26431 bp in length
* 181340 181439: gap of 100 bp
* 181440 195384: contig of 13945 bp in length
* 195385 195484: gap of 100 bp
* 195485 217207: contig of 21723 bp in length
* 217208 217307: gap of 100 bp
* 217308 218040: contig of 733 bp in length.

FEATURES          Location/Qualifiers
     source        1. 218040
                   /organism="Mus musculus"
                   /db_xref="taxon:10090"
                   /clone="RP23-153H17"
                   /clone_lib="RPCI-23 Female Mouse BAC"

     misc_feature  1. .7063
                   /note"assembly_fragment
                   clone_end:SP6
                   vector_side:left"
     misc_feature  7164. .8251
                   /note="assembly_fragment"
     misc_feature  8352. .9118
                   /note="assembly_fragment"
     misc_feature  9219. .10246
                   /note="assembly_fragment"

     misc_feature  10347. .53500
                   /note="assembly_fragment"
     misc_feature  53601. .55707
                   /note="assembly_fragment"
     misc_feature  55808. .59529
                   /note="assembly_fragment"
     misc_feature  59630. .61934
                   /note="assembly_fragment"
     misc_feature  62035. .66831
                   /note="assembly_fragment"
     misc_feature  66932. .72794
                   /note="assembly_fragment"
     misc_feature  72895. .77248
                   /note="assembly_fragment"
     misc_feature  77349. .84268
                   /note="assembly_fragment"
     misc_feature  84369. .93794
                   /note="assembly_fragment"
     misc_feature  93895. .107682
                   /note="assembly_fragment"
     misc_feature  107783. .125724
                   /note="assembly_fragment"
     misc_feature  125825. .137181
                   /note="assembly_fragment"
     misc_feature  137282. .154808
                   /note="assembly_fragment"
     misc_feature  154909. .181339
                   /note="assembly_fragment"
     misc_feature  181440. .195384
                   /note="assembly_fragment"
     misc_feature  195485. .217207
                   /note="assembly_fragment"
     misc_feature  217308. .218040
                   /note="assembly_fragment
                   clone_end:T7
                   vector_side:right".

BASE COUNT    60131 a    49124 c    47385 g    59388 t    2012 others
ORIGIN
```

```
SOURCE      Norway rat.
KEYWORDS    HTG; HTGS_PHASE1.
VERSION     AC114455.6  GI:21745639
ACCESSION   AC114455
DEFINITION  Rattus norvegicus clone CH230-249L20, *** SEQUENCING IN PROGRESS
            ***, 49 unordered pieces.
LOCUS       AC114455    162312 bp    DNA    linear    HTG 17-JUL-2002
AC114455/c
RESULT 5

Query Match  9.2%;  Score 192.6;  DB 2;  Length 218040;
Best Local Similarity  83.3%;  Pred No. 1.9e-42;
Matches  219;  Conservative  0;  Mismatches  44;  Indels  0;  Gaps  0;

Qy  799 ACAGGAACAAGCTTCAGCTCCGGGGCCTCCTTTACCTGCACACATGGCAGCCTGCTGTGTGTCT 858
        ||||||||| |||||| || |||||||||||||||||||||||| |||||| ||||||||| |||
Db 67800 ACAGGAGCAGCTTCGGCCATGGGTTCTTCCTCTCTTACCTGCCCACATGGGGCAGCCTGCTGTTTTTCT 67741

Qy  859 GGCGTGTGTCTTCGCCCATCTTTACTGCATCCATGTGGCCCCAAACACTTGGACACGTTATT 918
        || ||||||| |||| |||||||||| || |||||||||||| ||| ||||||| ||| | ||
Db 67740 GGATGTGTCCTGGCGAGTCTTTACGGCGGTCTATGTGGCCTTCAAGTACTTGGCCATTCTTGTG 67681

Qy  919 AACTCAGGGACAAAACCCTGGGAAAAACCATGAACCAATTGCCGCATGATATTTTATCTTCTTCTAGAA 978
        |||||||||| ||||| |||| ||||| ||||| || ||||| ||||||| || |||| |||| ||||
Db 67680 AACTCAGGGAAAAAAACTCCGGAGAAGCCATGGCCACTGGCATGATCCTTTATGTTTTACA 67621

Qy  979 ATATTTTTTCTTGTGTGCCTGGTGCACAGCTTTTTAAGTTTGTTCCCAGGGAGGTGTCTACGCTAGA 1038
        | || |||||| ||||| |||||||||||| ||||||||||||||||| |||||||||||||||||||
Db 67620 ACAITTTTCTGTGTCCCTGGTGCACAGCGTTTAAGTTTGTTCCCGGGAGGTGTCTACGCTAGA 67561

Qy 1039 GAAAGATCAGAIGTGCTTTTGGG 1061
        ||||||||| |||||||||||||
Db 67560 GAAAGATCTGATGTGCTTTTGGG 67538
```

ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
REFERENCE 1 (bases 1 to 162312)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
          Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
          Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
          Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
          Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
          Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
          Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
          Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
          Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
          Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
          Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
          Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
          Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
          Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
          Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
          Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
          Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
          Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
          Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
          Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
          Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
          Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
          Lozado,R.J., Lu,X., Lucier,R., Lucier,R., Luna,R., Ma,J.,
          Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
          Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
          Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
          Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
          Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
          Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
          Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
          Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
          Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
          Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,I.,
          Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
          Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
          Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
          Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
          Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
          Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
          Weinstock,G. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL Unpublished
REFERENCE 2 (bases 1 to 162312)
  AUTHORS Worley,K.C.
  TITLE   Direct Submission
  JOURNAL Submitted (09-MAR-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 162312)
  AUTHORS Worley,K.C.
  TITLE   Direct Submission
  JOURNAL Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          On Jul 14, 2002 this sequence version replaced gi:20467645.
COMMENT   Genome Center

          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu

            Project Information

          Center project name: GSXI
          Center clone name: CH230-249L20

            Summary Statistics

          Sequencing vector: plasmid;
          Chemistry: Dye-terminator Big Dye: 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 122635 bases at least Q40

          Consensus quality: 126261 bases at least Q30
          Consensus quality: 129845 bases at least Q20
          ----
          * NOTE: Estimated insert size may differ from sequence length
          * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 49 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *   1   1174: contig of 1174 bp in length
          *   1175   1274: gap of unknown length
          *   1275   2899: contig of 1625 bp in length
          *   2900   2999: gap of unknown length
          *   3000   4263: contig of 1264 bp in length
          *   4264   4363: gap of unknown length
          *   4364   5667: contig of 1304 bp in length
          *   5668   5767: gap of unknown length
          *   5768   7230: contig of 1463 bp in length
          *   7231   7330: gap of unknown length
          *   7331   8913: contig of 1583 bp in length
          *   8914   9013: gap of unknown length
          *   9014   10114: contig of 1101 bp in length
          *   10115   10214: gap of unknown length
          *   10215   11240: contig of 1026 bp in length
          *   11241   11340: gap of unknown length
          *   11341   12994: contig of 1654 bp in length
          *   12995   13094: gap of unknown length
          *   13095   14158: contig of 1064 bp in length
          *   14159   14258: gap of unknown length
          *   14259   15405: contig of 1147 bp in length
          *   15406   15505: gap of unknown length
          *   15506   17366: contig of 1861 bp in length
          *   17367   17466: gap of unknown length
          *   17467   18850: contig of 1384 bp in length
          *   18851   18950: gap of unknown length
          *   18951   21635: contig of 2685 bp in length
          *   21636   21735: gap of unknown length
          *   21736   22866: contig of 1131 bp in length
          *   22867   22966: gap of unknown length
          *   22967   24541: contig of 1575 bp in length
          *   24542   24641: gap of unknown length
          *   24642   26761: contig of 2120 bp in length
          *   26762   26861: gap of unknown length
          *   26862   28689: contig of 1828 bp in length
          *   28690   28789: gap of unknown length
          *   28790   31195: contig of 2406 bp in length
          *   31196   31295: gap of unknown length
          *   31296   34187: contig of 2892 bp in length
          *   34188   34287: gap of unknown length
          *   34288   37140: contig of 2853 bp in length
          *   37141   37240: gap of unknown length
          *   37241   38773: contig of 1533 bp in length
          *   38774   38873: gap of unknown length
          *   38874   41274: contig of 2401 bp in length
          *   41275   41374: gap of unknown length
          *   41375   42704: contig of 1330 bp in length
          *   42705   42804: gap of unknown length
          *   42805   44612: contig of 1808 bp in length
          *   44613   44712: gap of unknown length
          *   44713   47204: contig of 2492 bp in length
          *   47205   47304: gap of unknown length
          *   47305   49564: contig of 2260 bp in length
          *   49565   49664: gap of unknown length
          *   49665   52047: contig of 2383 bp in length
          *   52048   52147: gap of unknown length
          *   52148   55151: contig of 3004 bp in length
          *   55152   55251: gap of unknown length
          *   55252   58346: contig of 3095 bp in length
          *   58347   58446: gap of unknown length

```
LOCUS       AL158816             222885 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 6 clone RP3-405B4, *** SEQUENCING IN
            PROGRESS ***, 33 unordered pieces.
ACCESSION   AL158816
VERSION     AL158816.11  GI:9943990
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 222885)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 29, 2000 this sequence version replaced gi:9926474.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ405B4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 204390 bases at least Q40
            Consensus quality: 211527 bases at least Q30
            Consensus quality: 215491 bases at least Q20
            Insert size: 219685; sum-of-contigs
            Insert size: 126549; 14.0% error; agarose-fp
            Quality coverage: 3.63x in Q20 bases; sum-of-contigs Quality
            coverage: 6.72x in Q20 bases; agarose-fp
            --------------
            NOTE: This is a 'working draft' sequence. It currently
            consists of 33 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
            *
            *                1      3599: contig of 3599 bp in length
            *             3600      3699: gap of    100 bp
            *             3700     12818: contig of 9119 bp in length
            *            12819     12918: gap of    100 bp
            *            12919     16398: contig of 3480 bp in length
            *            16399     16498: gap of    100 bp
            *            16499     18590: contig of 2092 bp in length
            *            18591     18690: gap of    100 bp
            *            18691     22372: contig of 3682 bp in length
            *            22373     22472: gap of    100 bp
            *            22473     26905: contig of 4433 bp in length
            *            26906     27005: gap of    100 bp
            *            27006     29581: contig of 2576 bp in length
            *            29582     29681: gap of    100 bp
            *            29682     33446: contig of 3765 bp in length
            *            33447     33546: gap of    100 bp
            *            33547     48536: contig of 14990 bp in length
            *            48537     48636: gap of    100 bp
            *            48637     55599: contig of 6963 bp in length
            *            55600     55699: gap of    100 bp
            *            55700     65124: contig of 9425 bp in length
            *            65125     65224: gap of    100 bp
            *            65225     69090: contig of 3866 bp in length
            *            69091     69190: gap of    100 bp
            *            69191     73894: contig of 4704 bp in length
            *            73895     73994: gap of    100 bp
            *            73995     82665: contig of 8671 bp in length
            *            82666     82765: gap of    100 bp
            *            82766     87239: contig of 4474 bp in length
            *            87240     87339: gap of    100 bp
```

```
FEATURES             Location/Qualifiers
     source          1..162312
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-249L20"
BASE COUNT    40948 a  34779 c  33666 g  42210 t  10709 others
ORIGIN
            *            58447     61931: contig of 3485 bp in length
            *            61932     62031: gap of unknown length
            *            62032     65718: contig of 3687 bp in length
            *            65719     65818: gap of unknown length
            *            65819     69991: contig of 4173 bp in length
            *            69992     70091: gap of unknown length
            *            70092     73876: contig of 3785 bp in length
            *            73877     73976: gap of unknown length
            *            73977     77758: contig of 3782 bp in length
            *            77759     77858: gap of unknown length
            *            77859     80356: contig of 2498 bp in length
            *            80357     80456: gap of unknown length
            *            80457     86083: contig of 5627 bp in length
            *            86084     86183: gap of unknown length
            *            86184     90591: contig of 4408 bp in length
            *            90592     90691: gap of unknown length
            *            90692     94421: contig of 3730 bp in length
            *            94422     94521: gap of unknown length
            *            94522     99012: contig of 4491 bp in length
            *            99013     99112: gap of unknown length
            *            99113    102900: contig of 3788 bp in length
            *           102901    103000: gap of unknown length
            *           103001    106277: contig of 3277 bp in length
            *           106278    106377: gap of unknown length
            *           106378    111353: contig of 4976 bp in length
            *           111354    111453: gap of unknown length
            *           111454    118794: contig of 7341 bp in length
            *           118795    118894: gap of unknown length
            *           118895    124181: contig of 5287 bp in length
            *           124182    124281: gap of unknown length
            *           124282    133604: contig of 9323 bp in length
            *           133605    133704: gap of unknown length
            *           133705    141028: contig of 7324 bp in length
            *           141029    141128: gap of unknown length
            *           141129    149603: contig of 8475 bp in length
            *           149604    149703: gap of unknown length
            *           149704    162312: contig of 12609 bp in length.
```

```
RESULT 6
LOCUS       AL158816             222885 bp    DNA     linear   HTG 10-JUL-2001

Query Match   9.1%;  Score 191;  DB 2;  Length 163312;
Best Local Similarity 82.9%;
Matches 218;  Conservative 0;  Mismatches 45;  Indels 0;  Gaps 0;

Qy    799  ACAGGAACAGCTTCAGCTGCGGGGCGGGCTCCTTTACCTGCACACATGGGCAGCTGCTTGTGTCT  858
           ||||||  ||||||||| ||  |  | || |||||| ||||| || ||||||||| || |||| ||
Db  59064  ACAGGAGCAAACTTCAGCCATGGGTCTCTCCTTTATCTGCCCACCTGTGGGCAGCCGCCGTGTTTCT  59005

Qy    859  GGCTGCGTGTCTTCGCCATCTTTACTGCATCCATCGTGGCCCCAAACACTTGGACACCTTATT  918
           || ||||  | ||| |||| | | |    | || || ||||| || |||||||||||   ||| 
Db  59004  GGATGCGCACTGGCAGTCTTCACAGGGTCCATGTGGCCTCAAGTACTTGGACAGTCTTATG  58945

Qy    919  AACTCAGGGACAACAAACCTGGGAAAAACAATGACCATTGCCATGATPATTPTTATCTTCTAGAA  978
           ||||||||||||  || |||||| |  | |||||||| | ||||||  | ||  |||| ||||| ||
Db  58944  AACTCAGGGACAAAACCCTGGAGGAGGCCATGACCACTGGCATGAGCCTTTATGTTCTACAA  58885

Qy    979  ATATTTTTCTGTGTCCTGGTGCACAGCTTTTAAGTTGTCCCAGGAGGTCTCTACGGCTAGA  1038
           | ||||||| || |||||||||||||||||||||||| |  | |||||| | ||||| |||| |
Db  58884  GTGTTTTTTCTGTCCTGGTGCACAGCTTTTAAGTTGTTGTTCCTCGGAGGTGTCTACGCGCTAGA  58825

Qy   1039  GAAAGATCAGAGATGTGCTTTTGGG  1061
           ||||| | |  || |||||||||||
Db  58824  GAAAGGTCTGATGTGCTTTTGGG  58802
```

```
FEATURES             Location/Qualifiers
     source          1..222885
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-405B4"
                     /clone_lib="RPCI-3"
     misc_feature    1..3599
                     /note="assembly_fragment:00027"
     misc_feature    3700..12818
                     /note="assembly_fragment:00056"
     misc_feature    12919..16398
                     /note="assembly_fragment:00093"
     misc_feature    16499..18590
                     /note="assembly_fragment:00125"
     misc_feature    18691..22372
                     /note="assembly_fragment:00140"
     misc_feature    22473..26905
                     /note="assembly_fragment:00167"
     misc_feature    27006..29581
                     /note="assembly_fragment:00175"
     misc_feature    29682..33446
                     /note="assembly_fragment:00178"
     misc_feature    33547..48536
                     /note="assembly_fragment:00416"
     misc_feature    48637..55599
                     /note="assembly_fragment:00542"
     misc_feature    5700..65124
                     /note="assembly_fragment:00565"
     misc_feature    65225..69090
                     /note="assembly_fragment:00583"
     misc_feature    69191..73894
                     /note="assembly_fragment:00649"
     misc_feature    73995..82665
                     /note="assembly_fragment:00709"
     misc_feature    82766..87239
                     /note="assembly_fragment:00789"
     misc_feature    87340..93671
                     /note="assembly_fragment:00834"
     misc_feature    93772..97550
                     /note="assembly_fragment:00924"
     misc_feature    97651..100101
                     /note="assembly_fragment:01052"
     misc_feature    100202..103416
                     /note="assembly_fragment:01082"
     misc_feature    103517..107564
                     /note="assembly_fragment:01216"
     misc_feature    107665..131778
                     /note="assembly_fragment:01234"
     misc_feature    131879..135261
                     /note="assembly_fragment:01280"
     misc_feature    135362..137913
                     /note="assembly_fragment:01337"
     misc_feature    138014..140949
                     /note="assembly_fragment:01643"
     misc_feature    141050..156466
                     /note="assembly_fragment:01660"
     misc_feature    156567..164836
                     /note="assembly_fragment:01690"
     misc_feature    164937..168950
                     /note="assembly_fragment:01692.0"
     misc_feature    169051..173724
                     /note="assembly_fragment:01748"
     misc_feature    173825..179991
                     /note="assembly_fragment:02061"
     misc_feature    180092..190208
                     /note="assembly_fragment:02140"
     misc_feature    190309..193471
                     /note="assembly_fragment:02150"
     misc_feature    193572..197829
                     /note="assembly_fragment:02181"
     misc_feature    197930..222885
                     /note="assembly_fragment:00834"
```

```
*   87340   93671: contig of 6332 bp in length
*   93672   93771: gap of 100 bp in length
*   97551   97550: gap of 100 bp in length
*   97651  100101: contig of 3779 bp in length
*  100102  100201: gap of 100 bp in length
*  100202  103416: contig of 3215 bp in length
*  103517  103516: gap of 100 bp in length
*  103417  107564: contig of 4048 bp in length
*  107665  107664: gap of 100 bp in length
*  107565  131778: contig of 24114 bp in length
*  131879  131878: gap of 100 bp in length
*  131779  135261: contig of 3383 bp in length
*  135362  135361: gap of 100 bp in length
*  135262  137913: contig of 2552 bp in length
*  138014  138013: gap of 100 bp in length
*  137914  140949: contig of 2936 bp in length
*  141050  141049: gap of 100 bp in length
*  140950  156466: contig of 15417 bp in length
*  156567  156566: gap of 100 bp in length
*  156467  164836: contig of 8270 bp in length
*  164937  164936: gap of 100 bp in length
*  164837  168950: contig of 4014 bp in length
*  169051  169050: gap of 100 bp in length
*  168951  173724: contig of 4674 bp in length
*  173825  173824: gap of 100 bp in length
*  173725  179991: contig of 6167 bp in length
*  180092  180091: gap of 100 bp in length
*  179992  190208: contig of 10117 bp in length
*  190309  190308: gap of 100 bp in length
*  190209  193471: contig of 3163 bp in length
*  193572  193571: gap of 100 bp in length
*  193472  197829: contig of 4258 bp in length
*  197930  197929: gap of 100 bp in length
*  197830  222885: contig of 24956 bp in length.
```

```
BASE COUNT     69666 a    41364 c    44947 g    63682 t    3226 others
ORIGIN
                     /vector_side:right"
                     /clone_end:T7
```

```
RESULT 7
AL158816/c
LOCUS       AL158816      222885 bp    DNA    linear    HTG  10-JUL-2001
DEFINITION  Homo sapiens chromosome 6 clone RP3-405B4, *** SEQUENCING IN
            PROGRESS ***, 33 unordered pieces.
ACCESSION   AL158816
VERSION     AL158816.11  GI:9943990
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
Query Match             7.6%;  Score 160;  DB 2;  Length 222885;
Best Local Similarity  89.6%;  Pred. No. 2.8e-33;
Matches  172;  Conservative   0;  Mismatches  20;  Indels   0;  Gaps   0;

Qy    1863  CAGGTTGGGTTATGCAAGAAATCTCCCATGCTGAACGTGAGTGAGTTCAGAAATTCAGATGGC  1922
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157674  CAGGTTGGGTTATGCAAGAAATCTCCCATGCTGAACGTGAGTGAGTTCAGAAATTCAGATGGC  157733

Qy    1923  AAAATTTAGGATCCCTGAGTGACCCCCACTAATTAATTATATGAGAGACAACAGAAAAGTGGTCATACA  1982
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157734  AAAATTTAGGATCCCTGAGTGACCCCCACTAATTAATTATATGAGAGACAACAGAAAAGTGGTCATACA  157793

Qy    1983  CCACAGAGGAAGTTTCTGAGAAAAATTCATTTTAAATCCCAGATTTGGATCCTACAAAGAAGG  2042
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157794  CCACAGAGGAAGTTTCTGAGAAAAATTCATTTTAAATCCCAGATTTGGATCCTACAAAGAAGG  157853

Qy    2043  ACACAATTATGA  2054
            || ||||||
Db  157854  TTCCAGTTATGA  157865
```

```
REFERENCE   1  (bases 1 to 222885)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 29, 2000 this sequence version replaced gi:9926474.
            ---------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ---------- Project Information
            Center project name: dJ405B4
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 204390 bases at least Q40
            Consensus quality: 211527 bases at least Q30
            Consensus quality: 215491 bases at least Q20
            Consensus quality: ; sum-of-contigs
            Insert size: 219685; sum-of-contigs
            Insert size: 126549; 14.0% error; agarose-fp
            Quality coverage: 3.63x in Q20 bases; sum-of-contigs Quality
            coverage: 6.72x in Q20 bases; agarose-fp
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *       1   3599: contig of 3599 bp in length
            *    3600   3699: gap of 100 bp
            *    3700  12818: contig of 9119 bp in length
            *   12819  12918: gap of 100 bp
            *   12919  16398: contig of 3480 bp in length
            *   16399  16498: gap of 100 bp
            *   16499  18590: contig of 2092 bp in length
            *   18591  18690: gap of 100 bp
            *   18691  22372: contig of 3682 bp in length
            *   22373  22472: gap of 100 bp
            *   22473  26905: contig of 4433 bp in length
            *   26906  27005: gap of 100 bp
            *   27006  29581: contig of 2576 bp in length
            *   29582  29681: gap of 100 bp
            *   29682  33446: contig of 3765 bp in length
            *   33447  33546: gap of 100 bp
            *   33547  48536: contig of 14990 bp in length
            *   48537  48636: gap of 100 bp
            *   48637  55599: contig of 6963 bp in length
            *   55600  55699: gap of 100 bp
            *   55700  65124: contig of 9425 bp in length
            *   65125  65224: gap of 100 bp
            *   65225  69090: contig of 3866 bp in length
            *   69091  69190: gap of 100 bp
            *   69191  73894: contig of 4704 bp in length
            *   73895  73994: gap of 100 bp
            *   73995  82665: contig of 8671 bp in length
            *   82666  82765: gap of 100 bp
            *   82766  87239: contig of 4474 bp in length
            *   87240  87339: gap of 100 bp
            *   87340  93671: contig of 6332 bp in length
            *   93672  93771: gap of 100 bp
            *   93772  97550: contig of 3779 bp in length
            *   97551  97650: gap of 100 bp
            *   97651 100101: contig of 2451 bp in length
            *  100102 100201: gap of 100 bp
            *  100202 103416: contig of 3215 bp in length
            *  103417 103516: gap of 100 bp
            *  103517 107564: contig of 4048 bp in length
            *  107565 107664: gap of 100 bp
            *  107665 131778: contig of 24114 bp in length
            *  131779 131878: gap of 100 bp
            *  131879 135261: contig of 3383 bp in length
            *  135262 135361: gap of 100 bp
            *  135362 137913: contig of 2552 bp in length
            *  137914 138013: gap of 100 bp
            *  138014 140949: contig of 2936 bp in length
            *  140950 141049: gap of 100 bp
            *  141050 156466: contig of 15417 bp in length
            *  156467 156566: gap of 100 bp
            *  156567 164836: contig of 8270 bp in length
            *  164837 164936: gap of 100 bp
            *  164937 168950: contig of 4014 bp in length
            *  168951 169050: gap of 100 bp
            *  169051 173724: contig of 4674 bp in length
            *  173725 173824: gap of 100 bp
            *  173825 179991: contig of 6167 bp in length
            *  179992 180091: gap of 100 bp
            *  180092 190208: contig of 10117 bp in length
            *  190209 190308: gap of 100 bp
            *  190309 193471: contig of 3163 bp in length
            *  193472 193571: gap of 100 bp
            *  193572 197829: contig of 4258 bp in length
            *  197830 197929: gap of 100 bp
            *  197930 222885: contig of 24956 bp in length.
FEATURES             Location/Qualifiers
     source          1..222885
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-405B4"
                     /clone_lib="RPCI-3"
     misc_feature    1..3599
                     /note="assembly_fragment:00027"
     misc_feature    3700..12818
                     /note="assembly_fragment:00056"
     misc_feature    12919..16398
                     /note="assembly_fragment:00093"
     misc_feature    16499..18590
                     /note="assembly_fragment:00125"
     misc_feature    18691..22372
                     /note="assembly_fragment:00140"
     misc_feature    22473..26905
                     /note="assembly_fragment:00167"
     misc_feature    27006..29581
                     /note="assembly_fragment:00175"
     misc_feature    29682..33446
                     /note="assembly_fragment:00178"
     misc_feature    33547..48536
                     /note="assembly_fragment:00416"
     misc_feature    48637..55599
                     /note="assembly_fragment:00542"
     misc_feature    55700..65124
                     /note="assembly_fragment:00565"
     misc_feature    65225..69090
                     /note="assembly_fragment:00583"
     misc_feature    69191..73894
                     /note="assembly_fragment:00649"
     misc_feature    73995..82665
                     /note="assembly_fragment:00709"
     misc_feature    82766..87239
                     /note="assembly_fragment:00789"
     misc_feature    87340..93671
                     /note="assembly_fragment:00874"
     misc_feature    93772..97550
                     /note="assembly_fragment:00924"
     misc_feature    97651..100101
                     /note="assembly_fragment:01052"
     misc_feature    100202..103416
                     /note="assembly_fragment:01082"
     misc_feature    103517..107564
                     /note="assembly_fragment:01216"
```

```
        misc_feature    107665..131778
                        /note="assembly_fragment:01234"
        misc_feature    131879..135261
                        /note="assembly_fragment:01280"
        misc_feature    135362..137913
                        /note="assembly_fragment:01337"
        misc_feature    138014..140949
                        /note="assembly_fragment:01643"
        misc_feature    141050..156466
                        /note="assembly_fragment:01643"
        misc_feature    156567..164836
                        /note="assembly_fragment:01660"
        misc_feature    164937..168950
                        /note="assembly_fragment:01690"
        misc_feature    190309..193471
                        /note="assembly_fragment:01692.0"
        misc_feature    173825..179991
                        /note="assembly_fragment:01748"
        misc_feature    180092..190208
                        /note="assembly_fragment:02061"
        misc_feature    177930..222885
                        /note="assembly_fragment:02140"
        misc_feature    190309..193471
                        /note="assembly_fragment:02150"
        misc_feature    193572..197029
                        /note="assembly_fragment:02181"
        misc_feature
                        /note="assembly_fragment:00834"
                        /note="clone_end:T7"
                        /note="vector_side:right"

BASE COUNT     69666 a  41364 c  44947 g  63682 t   3226 others
ORIGIN

Query Match        7.2%;  Score 151;  DB 2;  Length 222885;
Best Local Similarity 100.0%;  Pred. No. 9.4e-31;
Matches 151;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1508 GGATTATGGCTTTCTCAAGATACCCAATTGTGAAANTCTGAGCATCACCTTCTTCCGTCAC 1567
             |||||||||||||| ||||||||||||||||||||| ||||||||||||||||||||||||
Db    215729 GGATTATGGCTTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCAC 215670

Qy      1568 CAGAGGGGGAGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAACGTGGTGG 1627
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    215669 CAGAGGGCGAGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAACGTGGTGG 215610

Cy      1628 ATTTTGTCCGTGACACACTTTGGGAACCACGA 1658
             |||||||| |||||||||||||||||||||||
Db    215609 ATTTTGTCGTGACACACTTTGGGAACCACGA 215579
```

RESULT 8

SCCHRIII/c

```
LOCUS       SCCHRIII      316613 bp    DNA     linear   PLN 29-JUN-2001
DEFINITION  S.cerevisiae chromosome III complete DNA sequence.
ACCESSION   X59720 S43845 S49180 S58084 S93798
VERSION     X59720.2  GI:1458895
KEYWORDS    .
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 316613)
  AUTHORS   Rad,M.R., Lutzenkirchen,K., Xu,G., Kleinhans,U. and Hollenberg,C.P.
  TITLE     The complete sequence of a 11,953 bp fragment from C1G on
            chromosome III encompasses four new open reading frames
  JOURNAL   Yeast 7 (5), 533-538 (1991)
  MEDLINE   91377317
   PUBMED   1897318
REFERENCE   2  (bases 1 to 316613)
  AUTHORS   Biteau,N., Fremeaux,C., Hebrard,S., Menara,A., Aigle,M. and
            Crouzet,M.
  TITLE     The complete sequence of a 10.8kb fragment to the right of the
            chromosome III centromere of Saccharomyces cerevisiae
  JOURNAL   Yeast 8 (1), 61-70 (1992)
  MEDLINE   92254505
   PUBMED   1580102
REFERENCE   3  (bases 1 to 316613)
  AUTHORS   Wilson,C., Bergantino,E., Lanfranchi,G., Valle,G., Carignani,G. and
            Frontali,L.
  TITLE     A putative serine/threonine protein kinase gene on chromosome III
            of Saccharomyces cerevisiae
  JOURNAL   Yeast 8 (1), 71-77 (1992)
  MEDLINE   92254506
   PUBMED   1580103
REFERENCE   4  (bases 1 to 316613)
  AUTHORS   Benit,P., Chanet,R., Fabre,F., Faye,G., Fukuhara,H. and Sor,F.
  TITLE     Sequence of the sup61-RAD18 region on chromosome III of
            Saccharomyces cerevisiae
  JOURNAL   Yeast 8 (2), 147-153 (1992)
  MEDLINE   92221691
   PUBMED   1561837
REFERENCE   5  (bases 1 to 316613)
  AUTHORS   Bolle,P.A., Gilliquet,V., Berben,G., Dumont,J. and Hilger,F.
  TITLE     The complete sequence of K3B, a 7.9 kb fragment between PGK1 and
            CRY1 on chromosome III, reveals the presence of seven open reading
            frames
  JOURNAL   Yeast 8 (3), 205-213 (1992)
  MEDLINE   92245758
   PUBMED   1574926
REFERENCE   6  (bases 1 to 316613)
  AUTHORS   Sor,F., Cheret,G., Fabre,F., Faye,G. and Fukuhara,H.
  TITLE     Sequence of the HMR region on chromosome III of Saccharomyces
            cerevisiae
  JOURNAL   Yeast 8 (3), 215-222 (1992)
  MEDLINE   92245759
   PUBMED   1574927
REFERENCE   7  (bases 1 to 316613)
  AUTHORS   Oliver,S.G., van der Aart,Q.J., Agostoni-Carbone,M.L., Aigle,M.,
            Alberghina,L., Alexandraki,D., Antoine,G., Anwar,R., Ballesta,J.P.,
            Benit,P. et al.
  TITLE     The complete DNA sequence of yeast chromosome III
  JOURNAL   Nature 357 (6373), 38-46 (1992)
  MEDLINE   92244356
   PUBMED   1574125
REFERENCE   8  (bases 1 to 316613)
  AUTHORS   Skala,J., Purnelle,B. and Goffeau,A.
  TITLE     The complete sequence of a 10.8 kb segment distal of SUF2 on the
            right arm of chromosome III from Saccharomyces cerevisiae reveals
            seven open reading frames including the RVS161, ADP1 and PGK genes
  JOURNAL   Yeast 8 (5), 409-417 (1992)
  MEDLINE   92327849
   PUBMED   1626432
REFERENCE   9  (bases 1 to 316613)
  AUTHORS   Wilson,C., Grisanti,P. and Frontali,L.
  TITLE     The complete sequence of a 6146 bp fragment of Saccharomyces
            cerevisiae chromosome III contains two new open reading frames
  JOURNAL   Yeast 8 (7), 569-575 (1992)
  MEDLINE   92397594
   PUBMED   1523889
REFERENCE   10  (bases 1 to 316613)
  AUTHORS   Scherens,B., Messenguy,F., Gigot,D. and Dubois,E.
  TITLE     The complete sequence of a 9,543 bp segment on the left arm of
            chromosome III reveals five open reading frames including
            glucokinase and the protein disulfide isomerase
  JOURNAL   Yeast 8 (7), 577-585 (1992)
  MEDLINE   92397595
   PUBMED   1523890
REFERENCE   11  (bases 26740 to 32076)
  AUTHORS   Defoor,E., Debrabandere,R., Keyers,B., Voet,M. and Volckaert,G.
  TITLE     Nucleotide sequence of D10B, a BamHI fragment on the small-ring
            chromosome III of Saccharomyces cerevisiae
  JOURNAL   Yeast 8 (8), 681-687 (1992)
  MEDLINE   93070606
   PUBMED   1441748
REFERENCE   12  (bases 169581 to 171116; 171683 to 172169)
  AUTHORS   Agostoni Carbone,M.L., Panzeri,L., Muzi Falconi,M., Carcano,C.,
            Plevani,P. and Lucchini,G.
  TITLE     Nucleotide sequence of 9.2 kb left of CRY1 on yeast chromosome III
```

JOURNAL   Yeast 8 (9), 805-812 (1992)
          analysis of open reading frame YCR28
          from strain AB972: evidence for a Ty insertion and functional
REFERENCE 13 (bases 1 to 315338)
  AUTHORS Valle,G.
  TITLE   TA-repeat microsatellites are closely associated with ARS consensus
           sequences in yeast chromosome III
  JOURNAL Yeast 9 (7), 753-759 (1993)
   MEDLINE 93377412
   PUBMED   8266725
          836809
REFERENCE 14 (bases 1 to 315338)
  AUTHORS Slonimski,P.P. and Brouillet,S.
  TITLE   A data-base of chromosome III of Saccharomyces cerevisiae
  JOURNAL Yeast 9 (9), 941-1029 (1993)
   MEDLINE 94091061
   PUBMED   1332309
REFERENCE 15 (bases 1 to 315339)
  AUTHORS Rodriguez-Cousino,N., Lill,R., Neupert,W and Court,D.A.
  TITLE   Identification and initial characterization of the cytosolic
           protein Ycr77p
  JOURNAL Yeast 11 (6), 581-585 (1995)
   MEDLINE 9373282
   PUBMED   7645349
REFERENCE 16 (bases 1 to 316613)
  AUTHORS MIPS.
  TITLE   Direct Submission
  JOURNAL Submitted (16-MAR-1992) MIPS, D-8033 Martinsried, FRG. Data
           collected by MIPS on behalf of the European Yeast Chromosome III
           Sequencing project
  REMARK   replaced by [13]
REFERENCE 17 (bases 1 to 314957)
  AUTHORS Jimenez,A.
  TITLE   Direct Submission
  JOURNAL Submitted (28-DEC-1992) MIPS, D-8033 Martinsried, FRG. Data
           collected by MIPS on behalf of the European Yeast Chromosome III
           Sequencing project. Update originating from A. Jimenez
  REMARK   replaced by [14]
REFERENCE 18 (bases 1 to 316613)
  AUTHORS Louis,E.J.
  TITLE   Direct Submission
  JOURNAL Submitted (25-JUN-1993) MIPS, D-8033 Martinsried, FRG. Data
           collected by MIPS on behalf of the European Yeast Chromosome III
           Sequencing project. Update originating from E.J. Louis
  REMARK   revised by [18]
REFERENCE 19 (bases 1 to 316613)
  AUTHORS Louis,E.J.
  TITLE   Direct Submission
  JOURNAL Submitted (14-FEB-1995) MIPS, D-8033 Martinsried, FRG. Data
           collected by MIPS on behalf of the European Yeast Chromosome III
           Sequencing project. Update originating from E.J. Louis
  REMARK   revised by [20]
REFERENCE 20 (bases 1 to 316613)
  AUTHORS Gromadka,R.
  TITLE   Direct Submission
  JOURNAL Submitted (29-JAN-1996) R. Gromadka, Protein Biosynthesis,
           Institute of Biochemistry and Biophysics, Pawinskiego 5A, Warsaw,
           Poland, Electronic Mail Address: robert@psd.ibb.waw.pl
  REMARK   revised by [21]
REFERENCE 21 (bases 1 to 316613)
  AUTHORS MIPS.
  TITLE   Direct Submission
  JOURNAL Submitted (17-JUN-2001) MIPS Yeast Genome Database, GSF -
           Ingolstaedter Landstrasse 1 D-85764 Neuherberg, Germany,
           mips-yeast-adm@gsf.de
  REMARK   Resequencing project, achieved by the joint effort of G. Valles and
           G. Volckaerts laboratories. Munich Information center for protein
           sequences.
COMMENT   On Jul 3, 2001 this sequence version replaced gi:1907116.
          The contig sequence which has been released to the file server in
          1992 has been subject to a resequencing project, achieved by the
          joint effort of G. Valles and G. Volckaerts laboratories. The

following lines present a summary of the altered entities. The
resequenced chrIII contig is 316613 bp with GGG-check: 429 GENETIC
ENTITIES valid no longer:
YCL01zw (YCL014w + YCL012w -> YCL014w / C_B1367 + C_C231->C_A1636),
PEP Y, ORF in identical form now C-terminal part of YCL014w. (A
frame shift in
the nt sequence of YCL014w leads to an elongated polypeptide.) This
elongated polypeptide encoded by the current YCL014w resembles the
Bud3p polypeptide from the
EMBL-Entry (Accession:SC175801.ID:U17580, total length 1636 aa)
YCL006c, PEP Y, ORF obsolete, is not an ORF in current Chromosome
III contig.
YCR062w (YCR061w -> YCR062w -> YCR061w / C_A583 + C_B120 -> C_B631),
PEP Y, ORF in identical form now C-terminal part of YCR061w.
YCR068w-a (YCR068w + YCR068w-a -> YCR068w / C_C429 + C_A145 -> C_
A520), PEP Y, ORF in identical form now C-terminal part of YCR068w.
YCR103c (C_F11), former ORF in region corresponding to coord
307255-307587(C)) is obsolete, open reading frame is destroyed by
single nt exchange resulting in a
stop codon after 45 aa.
NEW GENETIC ENTITIES:
YCL26c-b (C_E193(YCL027c-a)), PEP Y, new ORF
YCL021w-a (C_C125), PEP Y, new orf
ALTERED GENETIC ENTITIES:
YCL076w, PEP Y, peptide elongated N-terminally by 51 aa YCL074w,
PEP Y, 3 aa exchanged
YCL073c, PEP Y, 4 aa exchanged
YCL068c (C_F190->C_E260), PEP Y, peptide elongated N-terminally by
70 aa YCL065w, PEP Y, 1 aa exchanged
YCL064c, PEP Y, 3 aa exchanged
YCL061c (C_D853->C_F1096), PEP Y, peptide elongated N-terminally by

Query Match   6.3%;   Score 132.4; DB 8;   Length 316613;
Best Local Similarity   52.3%;   Pred. No. 1.6e-25;
Matches   342;   Conservative   0;   Mismatches 306;   Indels   6;   Gaps   2;

Qy   1281 GGTCTCTGCCCATCTGCCCTTTCAGGTTTGATATGACAATGAAAGGTGGTTCGTAGTCT 1340
           ||||| |||||||| ||| |||||| ||| ||||| |||||| || |||| ||| |
Db 145589 GATTACAGCTGGTATTTGGACTATCCACTTTGGTCTCGATAATGATATGTGGGCGATCTGA 145530

Qy   1341 AGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGATTTCATAACAATTTTGGGAGAG 1400
           ||| ||||| ||| || ||||| ||||||| |||| || || ||||| |||||
Db 145529 AGACAGAATGATCAAACCTTATTAAAGATATGGAACTAGATGTGGTTAGGTCTACTAGAAAC 145470

Qy   1401 TGATGCTTCTAAGCCCTATATGGGGAACACAATGACTTAACCATGTGGCTAGGGGAAAAGTT 1460
           ||||| ||| ||| | |||||||||||| | || ||| |||||| | ||||| || ||
Db 145469 AGATACACAAAAGAATTACCATGGGGAACACAGGGATCTAACTAGCAAACTAGCTCATGATTT 145410

Qy   1461 GGGTTTCTATACAGACACTTTGGTCCAAGCACACAAGGTATCACACTTGGGGGATTAGGCTTT 1520
           |||||| || ||| | ||||||||| |||||| || || | || ||| ||||||| ||
Db 145409 GAATATGTATGCAGATTTCGGACCAGGTCCAAATAAACATAACCATGGGGGCTGTGTTCTCT 145350

Qy   1521 GTCAAGATACCCAATTGTGAAAATCGAGCATCACCTTCTTCCGTCACCAGAGGCGAGAT 1580
           ||| ||||| |||| ||| |||| | | ||| || || | |||||| |||| |||
Db 145349 TTCTAAATTCCCTATCGTAAAATTCTACGCATCATTTATTGCCCCTCCAGTTGGGGAACT 145290

Qy   1581 CGCACCAGCCATCACATTGACCGTTAACATTTC---GGGCAAGCTGGTGGTGGATTTTGTCGT 1637
           |||||| ||||||| ||| | || |||| || || || | ||||| || ||||| || |
Db 145289 TGCGCGACGCCATTCATGCCACACTTCAAACGTACAACAATGACACTCGTTGACGGTCTTTGT 145230

Qy   1638 GACACACTTTGGGAACCACGAAGAATGACCTCGCACGAGAAAACTGCAGGCTATTGCTGTTTC 1697
           || | ||||||||| | ||| | || | || | | | || || || ||||| || | |
Db 145229 ATTCCATAGTGGACAAGAAGAGGATGAAGAGGATAGAAGACTGCAAAGTAACTACACATGGC 145170

Qy   1698 AAAACTACTGAAAAGTAGCTCTAATCAAGTGAATAGTTCTGGAGATATCACTTCAGCACC 1757
           || | || | || | | || || ||| || || || || | | |||| | ||| |||
Db 145169 TAAGCTCATGGACAATACGACTCCCCAGCTATTTATTAAGTTACTTAAGTGTTGTATCC 145110

Qy   1758 TGGCTCCAGAGATT---ATCTACAGCTCACTGAACATGCCAATGTGAAGGATATCGACAG 1814
           ||| | ||||||| | || | |||||| | | ||||| ||| || ||||| ||||
Db 145109 AGGTGAAGCGAACTACAATACGTACGTTAGTGAAACATCCGGAAATGCACGACAATTGATCC 145050

Qy   1815 CACTGATCATGACACAGATGGTGTGTGAATACATTATATTGTCAGGGCGGCTTGGGTTA 1874
           || | | ||| | |||| | ||| | | | ||||| | || |||||| |||||||||
Db   1698 CACTGATCATGACACAGATGGTGTGTGAATACATTATATTCGAGGGCGGCTTGGGTTA 1874

Db 145049 TTCTGACCATGATGAGATAGGTGTGTGAGTATATCTTGTTATAGGGCTTGAGAAGACAGGATA 144990
Qy 1875 TGCAAGAATCTCCATGCTGAACTGAGTGATTCAGAAAATTCAGATTGGCAAAATT 1928
           |||||||| |||| | ||| |||| ||||||| | ||||| |||| |||| ||| |||
Db 144989 TGCCTAGAGTGTGCAAGAGAGAAGAACGATAACCGATACGGACGCTACAAGTTGGTAAGTT 144936

RESULT 9
LOCUS   AC108828
DEFINITION   Mus musculus clone RP23-153H17, WORKING DRAFT SEQUENCE, 21 ordered
             pieces.
ACCESSION   AC108828
VERSION   AC108828.2 GI:20336119
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   Mus musculus
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 218040)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 218040)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
             Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
             Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,C.,
             Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
             Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
             Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
             Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
             Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
             Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
             Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
             Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
             McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
             Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
             Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
             Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
             Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
             Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
             Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
             Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
             Topham,K., Travis,N., Travis,N., Trigilio,J., Vassiliev,H.,
             Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
             Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (31-JAN-2002) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 218040)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
             Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
             Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
             Chazaro,C., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
             Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
             Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
             Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
             Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
             Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
             Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
             MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
             McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
             Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
             Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
             Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
             Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
             Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
             Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
             Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
             Topham,K., Travis,N., Travis,N., Trigilio,J., Vassiliev,H.,
             Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
             Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

LOCUS   AC108828   218040 bp   DNA   linear   HTG 28-APR-2002

TITLE   Direct Submission
JOURNAL   Submitted (28-APR-2002) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Apr 28, 2002 this sequence version replaced gi:18450089.
             All repeats were identified using RepeatMasker:
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             --------------- Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIBR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence_submissions@genome.wi.mit.edu
             --------------- Project Information
             Center project name: L20971
             Center clone name: 153_H_17
             --------------- Summary Statistics
             Sequencing vector: Plasmid; n/a; 100% of reads
             Chemistry: Dye-terminator Big Dye; 100% of reads
             Assembly program: Phrap; version 0.960731
             Consensus quality: 212760 bases at least Q40
             Consensus quality: 214579 bases at least Q30
             Consensus quality: 215279 bases at least Q20
             Insert size: 216040; sum-of-contigs
             Insert size: 220000; agarose-fp
             Quality coverage: 7.5 in Q20 bases; agarose-fp
             Quality coverage: 7.6 in Q20 bases; sum-of-contigs
             ----
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 21 contigs. Gaps between the contigs
             * are represented as runs of N. The order of the pieces
             * is believed to be correct as given, however the sizes
             * of the gaps between them are based on estimates that have
             * provided by the submittor.
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.
             *   1 7063: contig of 7063 bp in length
             *   7064 7163: gap of 100 bp
             *   7164 8251: contig of 1088 bp in length
             *   8252 8351: gap of 100 bp
             *   8352 9118: contig of 767 bp in length
             *   9119 9218: gap of 100 bp
             *   9219 10246: contig of 1028 bp in length
             *   10247 10346: gap of 100 bp
             *   10347 53500: contig of 43154 bp in length
             *   53501 53600: gap of 100 bp
             *   53601 55707: contig of 2107 bp in length
             *   55708 55807: gap of 100 bp
             *   55808 59529: contig of 3722 bp in length
             *   59530 59629: gap of 100 bp
             *   59630 61934: contig of 2305 bp in length
             *   61935 62034: gap of 100 bp
             *   62035 66831: contig of 4797 bp in length
             *   66832 66931: gap of 100 bp
             *   66932 72794: contig of 5863 bp in length
             *   72795 72894: gap of 100 bp
             *   72895 77248: contig of 4354 bp in length
             *   77249 77348: gap of 100 bp
             *   77349 84268: contig of 6920 bp in length
             *   84269 84368: gap of 100 bp
             *   84369 93794: contig of 9426 bp in length
             *   93795 93894: gap of 100 bp
             *   93895 107682: contig of 13788 bp in length
             *   107683 107782: gap of 100 bp
             *   107783 125724: contig of 17942 bp in length
             *   125725 125824: gap of 100 bp
             *   125825 137181: contig of 11357 bp in length
             *   137182 137281: gap of 100 bp
             *   137282 154808: contig of 17527 bp in length
             *   154809 154908: gap of 100 bp
             *   154909 181339: contig of 26431 bp in length
             *   181340 181439: gap of 100 bp
             *   181440 195384: contig of 13945 bp in length

```
FEATURES             Location/Qualifiers
source               1..218040
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-153H17"
        *            195385 195484: gap of  100 bp in length
        *            195485 217207: contig of 21723 bp in length
        *            217208 217307: gap of  100 bp
        *            217308 218040: contig of 733 bp in length.

misc_feature         1..7063
                     /clone_lib="RPCI-23 Female Mouse BAC"
                     /note="assembly_fragment"
                     vector_side:left"
                     clone_end:SP6
misc_feature         7164..8251
                     /note="assembly_fragment"
misc_feature         8352..9118
                     /note="assembly_fragment"
  ●     misc_feature 9219..10246
                     /note="assembly_fragment"
misc_feature         10347..53500
                     /note="assembly_fragment"
misc_feature         53601..55707
                     /note="assembly_fragment"
misc_feature         55808..59529
                     /note="assembly_fragment"
misc_feature         59630..61934
                     /note="assembly_fragment"
misc_feature         62035..66831
                     /note="assembly_fragment"
misc_feature         66932..72794
                     /note="assembly_fragment"
misc_feature         72895..77248
                     /note="assembly_fragment"
misc_feature         77349..84268
                     /note="assembly_fragment"
misc_feature         84369..93794
                     /note="assembly_fragment"
misc_feature         93895..107682
                     /note="assembly_fragment"
misc_feature         10783..125724
                     /note="assembly_fragment"
misc_feature         125825..137181
                     /note="assembly_fragment"
  ●     misc_feature 137282..154808
                     /note="assembly_fragment"
misc_feature         154909..181339
                     /note="assembly_fragment"
misc_feature         18140..195384
                     /note="assembly_fragment"
misc_feature         195485..217207
                     /note="assembly_fragment"
misc_feature         217308..218040
                     /note="assembly_fragment"
                     vector_side:right"
                     clone_end:T7

BASE COUNT    60131 a  49124 c  47385 g  59388 t   2012 others
ORIGIN
```

```
Query Match          5.8%;     Score 122.6; DB 2;  Length 218040;
Best Local Similarity     75.6%;   Pred. No. 8.9e-23;
Matches 152; Conservative     0; Mismatches 49; Indels  0; Gaps  0;

Qy       38 TGCTGGGATGTGTTTCTGGGTCTCTTACGATGACCTGGGACCGACGGATGATCATTACTTTC  97
            ||| | ||||||||||||| |||||||||||||||||||| ||| || ||||| |||  |||
Db    88151 TCCTAGGCTACGGTCTCCCTGGCTCTCTACCACGGACTGAGCCCCATGATCTATTATTTTTC 88210

Qy       98 CTTTGCAAACACTAGAACTCACTGGGCTTGAAGGTTTTAGTAGTTAAGCAATTTCTTTCTCCAA 157
            ||||  || ||||||||||||||||| |||||||| | |||||| |||||||| | ||||||
Db    88211 CCCTGCAAAACGCTGGGAGCTCACAGGGCCTGGGAATTCTTTTTGTGTGTTAGCCTTTCTCTCCGA 88270

Qy      158 TATTCCTAACAATTACTCCTTTCTGGGAAATTGGTTAACAAAGAAGTGGGATGCTAACCCTGC 217
            ||| |||||||||| |||||| | || |||| || || ||  |||| |||||||||| | | |||
```

```
Db       88271 TACTCTTAACAATTCCCCACTCTTGGGAAACTCGTCAATAAGAAGTGGACACTCTTCTCTGC 88330
Qy         218 TGAGGATAATCACTATTGGCA 238
               ||||||  |||||||||||| |||
Db       88331 TGAGGATCGTCACTGTAGGTA 88351
```

```
RESULT 10
LOCUS       D89130              1234 bp    mRNA    linear    PLN 13-MAR-1998
DEFINITION  Schizosaccharomyces pombe mRNA, partial cds, clone: SY 0503.
ACCESSION   D89130
VERSION     D89130.1  GI:1749467
KEYWORDS    .
SOURCE      Schizosaccharomyces pombe (strain:PR745) cDNA to mRNA,
            clone_lib:library of H. Nojima clone:SY 0503.
ORGANISM    Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (sites)
AUTHORS     Yoshioka,S., Kato,K., Nakai,K., Okayama,H. and Nojima,H.
TITLE       Identification of open reading frames in Schizosaccharomyces pombe
            cDNAs.
JOURNAL     DNA Res. 4 (6), 363-369 (1997)
MEDLINE     98162722
REFERENCE   2  (bases 1 to 1234)
AUTHORS     Yoshioka,S.
TITLE       Direct Submission
JOURNAL     Submitted (15-NOV-1996) Sachiyo Yoshioka, Tsukita Cell Axis Project
            ERATO JST, Kyoto Research Park; 17 Chudouji Minamimachi,
            Shimokyo-ku, Kyoto, Kyoto 600, Japan
            (E-mail:syoshi@cell.tsukita.jst.go.jp, Tel:+81-75-315-7913,
            Fax:+81-75-315-6420)
FEATURES             Location/Qualifiers
source               1..1234
                     /organism="schizosaccharomyces pombe"
                     /strain="PR745"
                     /db_xref="taxon:4896"
                     /clone="SY 0503"
                     /clone_lib="library of H. Nojima"
CDS                  <1..1207
                     /note="similar to Saccharomyce cerevisiae hypothetical
                     107.9KD protein in POL4-SRD1 intergenic region, SWISS-PROT
                     Accession Number P25618"
                     /codon_start=2
                     /protein_id="BAA13792.1"
                     /db_xref="GI:1749468"
                     /translation="IYSFASIRISSFSPAKVWGGAFLVYILYSLAHWWVAYEFVPG
                     GPILRERTSYILIFIGWNLAALVPAYSCESKEPNKADSSVVDIKQSDSSYRRSFKKS
                     LLTGFCLALMALKFAIQNMPPYDYTPYHNEKLFTAGIWTHIHFGLDNFMYASENRIRD
                     AVRDMELDVFGLLESDYDTQRLIMFRDLTQVLAHDLGMYADYGPGDKHFWGAALLSKF
                     PIVNSTHHLLPSPQGELAPAIHATLDVYGELIDVVVSHNGQYESOLDRRLQSTELARI
                     MRESPRPLVFLGIVVSNVGQEPQTILTRDTGMLDIEPADYDRWCQYIFYRGVKRIGYA
                     RLHRSTIIDTELQTDGKFLVTKDLGRNVRIDKEHVPESHRYPSLFEGTGVNGHYDDNL
                     VVHEPWYD"
BASE COUNT      333 a    253 c    253 g    395 t
ORIGIN
```

```
Query Match          5.8%;     Score 121.8; DB 8;  Length 1234;
Best Local Similarity     49.0%;   Pred. No. 9.1e-23;
Matches 324; Conservative     0; Mismatches 337; Indels  0; Gaps  0;

Qy     1283 TCTCTGCTCCACCATCTGGCCTTTCAGGTTTGTGGATATGACAAATGAAGGTGGTCTAGTCTAG 1342
            || |||||| || |||||| ||||||| ||| | | | || ||||||| || |||| || |
Db      408 TCACTGCCTCGTGTATTTGGACTATTCACTTTCGGCTTAGACAATTTCATGTATGCAAGGTGAGA 467

Qy       98 AAAGATCAGCTCACCTGCTCCAATCAAAAACAGGTGCAGATTTCATAACAATTTTGGAGAGTG 1402
            || ||| | || | |||  |  || || || ||||| || || || | ||| | | ||||
Db      468 ATCGTATACGCGATGCTGTTCGAGATAGTGGGAAGATGTGGGATGTCTGTTTGGTCTGAATCG 527

Qy     1403 ATGCTTCTAAGCCCTATATGGGGGGACAATGAACTTAAACCATAGTGTGGCTAGGGGAAAGTGG 1462
            ||||||
```

```
RESULT 11

Qy    528 ACACAACAACGTTTAATTATATGGGATTTCCGAGATTTAACCACGGGTTTTACCACATGATCTAG 587
Db   1463 GTTTCTATACAGACTTTGGTCCCAAGCACAAGGTATCACACTTGGGGGATTATGCGTTTGT 1522

Db    588 GAATGTATGCCGGATTATGCGCCCAGGCCCTGATAAACATACTTGGGCGGCCTGCCTTACTCT 647
Qy   1523 CAAGATACCCAATTGTGAAAATCTGAGACATCACCTTCTTCCGTCACCAGAGGGCGGAGATCG 1582

Db    648 CCAAATTTCTTATTGTCAAACTCGACTCACCATTGTTACCTTCTTCCTCAAGGAGAACTTG 707
Qy   1583 CACCAGGCCATCACATTGACCGTTAACATTTCGGGCAACCTGTGGGATTTTGTCGTGACAC 1642

Db    708 CTCCCTGCTATTCATTGCAACACTAGAATGTCTACGGAGAGCCTAATAGATAGATGTTTGTCTC 767
Qy   1643 ACTTTGGGAACCACACGAAGATGAGACCTCGACACAGGAAAACTGCCTATTGCCTGGTTTCAAAC 1702

Db    768 ACAAATGGTCAAATAGAAAGTCAGTTGGATCGACGTTCTTCAGACCACTCAGTTCGGCTCGAA 827
Qy   1703 TACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGGATAATATCACTTCAGCACCTGGCT 1762

Db    828 TTATTGCGCGGAGAGTCCAAGGCCTCTTAGTGTTTCTCGGGATAATGTTTGTTTCAAACGTTGGCC 887
Qy   1763 CCAAGAATTATATCACAGCTCACTGAACATGGCAATGTGAAGGATAATGTGACACGACACTGATC 1822

Db    888 AAGAACCCCAAAACTATTCTAAGCGCAGACACGGGAATGTTGTAGACATTTGAAACCTGCTGACT 947
Qy   1823 ATGACAGATGGTTGTTGAATAACATTAATGTTATCGCAGAGTGGCGTTATGCAAGAA 1882

Db    948 ATGATCGTTGGTTGTCAGTATATAATTTTATCGAGGTGTCAAGCGAATTGGTTATGCTAGAC 1007
Qy   1883 TCTCCCATGCTGAAACTGAGTGATTCAGAAAATTTAGGATCCCTGATG 1942

Db   1008 TTCATCGCCTCTACTATAACAGACACCGAGCTTCAAACCGAAAAGTTTTTGGTTTACCAAAG 1067
Qy   1943 —                                                          1943

Db   1068 —                                                          1068
```

genome is now being continued with funding from The European
Commission. Fourteen European sequencing laboratories, including
the Sanger Centre, are participating in the project.
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program SpSplice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites.
CDS are numbered using the following system eg SPBC25H2.01c SP (S.
pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c
(complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid c688 is overlapped at the 3' end by cosmid c3G9, EMBL entry
SPAC3G9, accession number AL021046.

```
FEATURES             Location/Qualifiers
     source          1..34034
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /map="I"
                     /clone="cosmid c688"
     gene            1..1280
                     /gene="SPAC688.01"
     CDS             1..1280
                     /gene="SPAC688.01"
                     /partial
                     /gene="SPAC688.01", len:425, SIMILARITY:Saccharomyces
                     cerevisiae, YCQ7_YEAST, hypothetical 107.9 kd protein.,
                     (953 aa), fasta scores: opt: 1034, E():0, (46.5% identity
                     in 430 aa)"
                     /codon_start=3
                     /label=SPAC688.01
                     /product="hypothetical protein similar to yeast YCR017C
                     protein involved in sensitivity to certain drugs"
                     /protein_id="CAB90768.1"
                     /db_xref="GI:7768477"
                     /translation="SYFMYINHGWCSYLGGLLIFTSYVLIYSFASIRISSFYSPAKVNG
                     GAFLVYILYSLAHVWAVAYEFVPGGPILRERTSYILIFIGWNLAALVPAYSCESKEPN
                     KADSSVVDIKQSODSSYRRRSFRFKKSLLTGFCLALMALKFAIONMPPYDYTPYHPNEKLF
                     TAGIWTIHFGLDNFMYASENRIRDAVRDMELDVFGLLESDTDLRLIMGERDLTQVLAHD
                     LGMYADVGPGPDKHTWGAALLSKFPIVNSTHHLLPSPQGELAPAIHATLDVYGELIDV
                     VSHNGQYESOLDRRLQSTELARIMRESPRPLVFLGYVVSNVGQOEPQTILTRDTGMLD
                     IEPADYDRWCQVIFYRGVKRICYARLIRSTIITDTELQTGKFLVTKDLGRNVRIDKEHV
                     PESHRYPSLFEGTGVNGHYYDNLVYHEPWYD"
                     /complement(join(1545..1979,2109..2256,2343..2392))
     gene            complement(join(1545..1979,2109..2256,2343..2392))
                     /gene="SPAC688.02c"
     CDS             complement(join(1545..1979,2109..2256,2343..2392))
                     /gene="SPAC688.02c"
                     /note="SPAC688.02c, len:210, chromosome xvi reading frame orf
                     cerevisiae, Q12143, hypothetical protein with very low similarity to
                     ypl233w., (216 aa), fasta scores: opt: 183, E():9.1e-05,
                     (22.7% identity in 185 aa)"
                     /codon_start=1
                     /label=SPAC688.02c
                     /product="hypothetical protein with very low similarity to
                     yeast ypl233w"
                     /protein_id="CAB90769.1"
                     /db_xref="GI:7768478"
                     /translation="MSSHGDASVLPKIALESINDYSVVKSVPRQAIOEKLAIHLPEQA
                     SREAGMGREDQLRIRVKEMLNELIESTFSIIQENVTINGFDANSALQDPKNSDQIEPF
                     DLALRTRVQOLENEVEDAHVLVARYRKSVPAQYEKAVVDAMEQOTAFLRNVKDDYSL
                     ODKEVENPDEQYSTTVFRKEDIDRYEQTIAKVAYLKKNLPRVVARLEKTP"
                     /complement(1980..1990)
                     /gene="SPAC688.02c"
     misc_feature    complement(1980..1990)
                     /gene="SPAC688.02c"
```

```
misc_feature    complement(2103..2108)
                /gene="SPAC688.02c"
                /note="ttaacatatag, splice branch and acceptor"

misc_feature    complement(2257..2271)
                /gene="SPAC688.02c"
                /note="gtgtgt, splice donor sequence"

misc_feature    complement(2337..2342)
                /gene="SPAC688.02c"
                /note="ttaacactttatag, splice branch and acceptor"

gene            complement(join(2596..2953,3083..3297,3381..3422))
                /gene="SPAC688.03c"

CDS             complement(join(2596..2953,3083..3297,3381..3422))
                /gene="SPAC688.03c"
                /note="SPAC688.03c, len:204, SIMILARITY:Homo sapiens,
                Q9Y4X0, ammecr1 protein, (333 aa), fasta scores: opt:
                443, E():3.4e-25, (40.3% identity in 186 aa)"
                /codon_start=1
                /label=SPAC688.03c
                /product="conserved hypothetical protein"
                /protein_id="CAB90770.1"
                /db_xref="GI:7768479"
                /translation="MICWPSLNKESWNDKEYCYCFEVVAATLEHRKVRDKWNAKSWT
                RSIPLFVKFASGKGHDKQLRGCGIGTFRARPLVTNLTYFSKQAAFCDERFRPISLGELA
                LLECQIDLLVDFEPIDDPLDWEVGITHGVSIKFTANGIRYSSTYLPSVAAEQRWDOEET
                LESLIHKAGYYGSIRSLQTITATRYKSLEICGCTYEEYLHNLELLG"

misc_feature    complement(join(2662..2953,3083..3273))
                /gene="SPAC688.03c"

misc_feature    complement(2954..2976)
                /gene="SPAC688.03c"
                /note="gtaagt, splice donor sequence"

misc_feature    complement(3298..3315)
                /gene="SPAC688.03c"
                /note="ctaatacaacatttaaaaatag, splice branch and
                acceptor"

misc_feature    complement(3077..3082)
                /gene="SPAC688.03c"
                /note="gtaagt, splice donor sequence"

misc_feature    complement(3375..3380)
                /gene="SPAC688.03c"
                /note="gtatga, splice donor sequence"

misc_feature    complement(3375..3380)
                /gene="SPAC688.03c"
                /note="ctaatacaaatggacaag, splice branch and acceptor"

misc_feature    complement(4047..4775)
                /gene="SPAC688.04c"

gene            complement(4047..4775)
                /gene="SPAC688.04c"

CDS             complement(4047..4775)
                /gene="SPAC688.04c"
                /note="SPAC688.04c, len:242, SIMILARITY:Saccharomyces
                cerevisiae, YIV8_YEAST, Glutathione transferase GTT1, (234
                aa), fasta scores: opt: 274, E():3.5e-11, (27.0% identity
                in 222 aa)"
                /codon_start=1
                /label=SPAC688.04c
                /product="putative glutathione S transferase"
                /protein_id="CAB90771.1"
                /db_xref="GI:7768480"
                /translation="MIVLHHLKNSRSTRIVWMLEELKVPYEIKVYDRVDGRAPPAYTK
                LSPLGKSPIVVDDGVTYIESAAILEHLVRKYGPSEPKPSEEDVAELEKYELWMHFSEAS
                LMPFIWASHVLDLSVNMTPIFFRYIVROYVNGIKSKYLSKETFLNLDYIDNHLASNEY
                FAGEQFTAADPQMCFPIFAAQRDYLSQKPYKNIKRMWRVVSDRPACRIAEKVEDNTL
                TLFSDVERYSHPPTPPPEQVRSDE"

misc_feature    complement(4164..4775)
                /gene="SPAC688.04c"
                /note="Match to PF00043 GST, Glutathione S-transferases.
                Score 85.13"

gene            complement(join(5637..6842,6927..6980))
                /gene="SPAC688.06c"

CDS             complement(join(5637..6842,6927..6980))
                /gene="SPAC688.06c"
                /note="SPAC688.06c, len:419, predicted to contain
                coiled-coil region, SIMILARITY:S. cerevisiae, LOW to
                YB61_YEAST, (324 aa), Actin-related protein, regulator of
                microtubule stability,
```

```
                (21.3% identity in 300 aa)"
                /codon_start=1
                /label=SPAC688.06c
                /product="hypothetical protein very putative Actin-related
                protein, regulator of microtubule stability"
                /protein_id="CAB90772.1"
                /db_xref="GI:7768481"
                /translation="MSAEEYIEVSSSPDIFIDDDDMITIEPELNKNPKDCNSKRKRSV
                TECCEIRLHTSKCDFESTOOLVHHNCTGHKVHEHNLNAVDEDFDTENLPLLFSSFSD
                NESDIIEPDLNTRVAADNDVLLSRYKIKINASASCRRNTFEHSAYHSREEISSSGFYYH
                RKPQLFEKSLERKLGNKSIEANRSPLIKELCEEANSTENVCFSVSTVDEIQORHPSAGH
                SIDSTCQSNSFELEGDSATHKKKTDNIKEFTSCEFNDRSRTLLNYAGYMDTNKNADNE
                AKSLKLRKLENPVEKLRAIAESYGFKSSDSKATLIKIVESCLDAIDSRSQSRKKLGKET
                PHDYLITSTKVLEFDIIVTQTHRAISQVVKQAKDNSVWIKILTYSAIDVEEFQLWLK
                RKNLNVSLDLIKSWCDKYGVLMKGSWH"

gene            complement(6843..6857)
                /gene="SPAC688.06c"

misc_feature    complement(6921..6926)
                /gene="SPAC688.06c"
                /note="ttaactgagctttag, splice branch and acceptor"

misc_feature    complement(7359..10475)
                /gene="SPAC688.06c"
                /note="gtatga, splice donor sequence"
```

```
Query Match    5.8%;  Score 121.8; DB 8; Length 34034;
Best Local Similarity 49.0%;  Pred. No. 1.3e-22;
Matches  324; Conservative  0; Mismatches 337;  Indels  0; Gaps  0;

Qy    1283 TCTCTGCTTGCCCATCTGCCCTTCAGGTTTGGATATGACAAATGAAGGTGGTCTAGTCTAG 1342
                |  |||  |  |  |||  ||  ||    |||||| |  || | |  |||||  ||  ||||
Db     481 TCACTGCTGTGATTTGGACTATTCACTTTGGCTTAGACAAATTTCATGTATGCAAGTGAGA  540

Qy    1343 AAAGATCAGCTCACCTGCTTCAAATGAAACAGGTCGCAGAATTCATTGATGACAATTTGG 1402
                ||  |||    | |  ||  || |||| |  ||  ||| ||| |  ||| ||  |  |||
Db     541 ATCGTATACGCGATGCGTGTTCGAGATATTGGAAACTTGATGTCTTTGGTGCTACTGGCTTTGT 600

Qy    1403 ATGCTTCTAAGCCCTATATGTGGGAACAATGACTTAACCATGTGGCTAGGGAAAGTTGG 1462
                || |||  ||  |||| | |||| |  |  ||  ||| | |  ||||| |  ||  || ||
Db     601 ACACACAACGTTTAAATTATGGGATTTCGAGATTTTGGGACTTTTAGGCACACATGATCTAG 660

Qy    1463 GTTTCTATACAGACTTTGGTCCAAGCACAAAGGTATCACACTTGGGGGATTATGGCTTTGT 1522
                |||  | |  ||  |||  || ||   | ||  |   |  ||  |  |||  ||  |||| |
Db     661 GAATGTATGCGGATTATGGCCCAGGCCCTGATAAACATACTTGGGGCCGCGGCTTTACTCT 720

Qy    1523 CAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTGACCAGAGGGGGAGATCG 1582
                | || || |  | |  ||| |  ||  ||| |  ||  ||  |  || |  | | |  |||
Db     721 CCAAATTTCCTATTGTCAACTCGACTCACCATTTGTTACCTTCCTCCCTCCAAGGAGAACTTG 780

Qy    1583 CACCAGCCATCACATTGACCGTTAACATTTCGGCAACGTCGGTGGATTTTGTCGTGACAC 1642
                | || |  ||| || || |   || | |||  |  | |  | ||  ||| ||  | |||
Db     781 CTCCTGCTTATTCATTGCAACGACTAGAATGTTCTACGGAGCTAATAGAATGTTTGTTCTC 840

Qy    1643 ACTTTGGGAACCACGAAGATGACCTCGACAGGAGAAACTCGAGGCTTATTGCTGTTTCAAAAC 1702
                | || ||  ||| || ||| |  |  ||  ||| | | |  |||  || |  |||  | ||
Db     841 ACAATGGTCAATATGAAAGTCAGTTGGATCGACGTCTTCAGAGCGACTGAGTTGGCTCGAA 900

Qy    1703 TACTGAAAAGTAGCTCTAATCAAGTGATATATTTCTGGGATATAATCACTTCAGCACTGGCT 1762
                || ||   ||   |||  ||| |   | || |||| |  || | |  |||  | || | |||
Db     901 TTATGCCCGGAGACTCCAAGGCCCTCTAGTGTTTTCTCGGAATAATGTTGTTCAAACGTTGGCC 960

Qy    1763 CCAGAGATTATCTCACAGCCTCACTGAACATGGCAATGTGAAGGAGATATCGACAGCACTGATC 1822
                || | |  |  |||| |  | | ||||| |  |  ||  ||  | | |  ||  || |   |  |
Db     961 AAGAACCCCAAACTATTCTAACGCGAGACACGGGAATGTTAGACATTGAAACCTGCTAGU 1020

Qy    1823 ATGACAGATGGGTCGGAATAACATTATTGTATATCGAGGGCGTGATCACGTTGGGTTATGCAAGAA 1882
                ||| | | || ||  |  ||||| | |  |  |  || ||  |  ||||  ||  |  | | |
Db    1021 ATGATCGTTGGTTGTCAGTATAATTTTATTTAATCGAGGTGTCAAGCGAATTGGTTATGCTAGAC 1080

Qy    1883 TCTCCCATGCTGAACTGAGTGATTCAGAAAATTAGAGATCCCTGATG 1942
                ||   |  | | |  |||   ||  |  | |||| | | | | |
Db    1081 TTCATCGCTCTACTATTAACAGACACCCGAGCTTCAAACCGGAAAGTTTTTGGTTACCAAAG 1140

Qy    1943 A 1943
```

Db   1141 A 1141

RESULT 12

LOCUS   AX337478/c
DEFINITION Sequence 7987 from Patent WO0194629.
ACCESSION   AX337478
VERSION   AX337478.1 GI:18128197
KEYWORDS
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
TITLE   Cancer gene determination and therapeutic screening using signature
            gene sets
JOURNAL   Patent: WO 0194629-A 7987 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES   Location/Qualifiers
source   1..305
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT   81 a   45 c   44 g   127 t   8 others
ORIGIN

Query Match 5.3%; Score 112; DB 6; Length 305;
Best Local Similarity 94.1%; Conservative 0; Pred. No. 4.5e-20;
Matches 112; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy   1985 ACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTTACAAAGAAGGAC 2044
          ||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db   305 ACNGAGAAGTTTCTGNGAAAAATTCATTTTAATCCCNGATTTGGATCCTTACAAAGAAGGNC 246

Qy   2045 ACAATTATGAAAAACAACCATCATTTTCATATGAATACTCCCAAATACTTTTTTATGAAAC 2103
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||| ||||||
Db   245 ACAATTATGAAAACAACCACCATCATTNTCATATGAATACTCCCAAATACTTNTNTNTGAAAC 187

RESULT 13

LOCUS   CNS06JWV
DEFINITION T3 end of clone XA50AA002B08 of library XA50AA from strain CLIB 533
            of Saccharomyces bayanus, sequence tagged site.
ACCESSION   AL402149
VERSION   AL402149.1 GI:12160780
KEYWORDS   STS.
SOURCE   Saccharomyces bayanus.
ORGANISM   Saccharomyces bayanus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1 (bases 1 to 995)
AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuvegliso,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
TITLE   Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE   20584711
PUBMED   11152876
REFERENCE   2 (bases 1 to 995)
AUTHORS   Bon,E., Neuvegliso,C., Casaregola,S., Artiguenave,F., Wincker,P.,
            Aigle,M. and Durrens,P.
TITLE   Genomic exploration of the hemiascomycetous yeasts: 5.
            Saccharomyces bayanus var. uvarum
JOURNAL   FEBS Lett. 487 (1), 37-41 (2000)
MEDLINE   20584715
PUBMED   11152880

Db   1141 A 1141

REFERENCE   3 (bases 1 to 995)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT   This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES   Location/Qualifiers
source   1..995
            /organism="Saccharomyces bayanus"
            /strain="CLIB 533"
            /variety="uvarum"
            /db_xref="taxon:4931"
            /clone="XA50AA002B08"
            /clone_lib="XA50AA"
            /note="end : T3"
misc_feature <3..>958
            /note="similar to Saccharomyces cerevisiae ORF YCR017c [
            similarity to hypothetical S.pombe protein]
            1 putative frameshift(s)"
            /evidence=not_experimental
BASE COUNT   276 a   190 c   210 g   318 t   1 others
ORIGIN

Query Match 5.3%; Score 110.6; DB 11; Length 995;
Best Local Similarity 52.2%; Pred. No. 1.3e-19;
Matches 314; Conservative 1; Mismatches 280; Indels 6; Gaps 3;

Qy   1277 AAGAGGTCTCTGCTGCCAATCTGGCCTTTCAGGTTTTGGATATGACAAATGAAGGGTGGTCTA 1336
          |||||||||| | ||| |||| || || || | |||||||| || ||||| || |||| ||
Db   391 AATTGATCTCTGCTGGTATTTGGACCACTCCATTTTGGTCTTGATAAATGATATGTGGGCAT 450

Qy   1337 GTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGATTTCATACAACAATTTGG 1396
          | | |||| ||| ||| | || | |||| ||||||| || || |||||| || |||||
Db   451 CTGAAGCAGAATGATCAACTTAATAAAAAGARAAATTGGAGTTGGAGTTGGTAGGATTACTAG 510

Qy   1397 AGAGTGATGCTTCTAAGCCCCTATATGGGGAACAATGACTTAACCATGTGGCTAGGGGAAA 1456
          | || |||| | | ||| | || || |||| ||||| ||| ||||| || || ||||
Db   511 AAACAGATACGCAAAGAAATTATTATGGGTAACAGGGATCTAACCAGTAAACTAGCCCATG 570

Qy   1457 AGTTGGGTTTCTTATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGGGGGATTATGG 1516
          || |||| || | || ||| ||| || || | |||||| ||||| |||| || || ||
Db   571 ATCTGAAATAGTATGCTGAATTTGGCCCCGGGTCCAAAACAAACACACATGGGGGTCTGTATTC 630

Qy   1517 CTTTGTCAAGATACCCAATTGTGAAAATCTGAGCAATCACCTTCTTCCGTCACCAGAGGGCG 1576
          | ||| ||||| | || || || || || ||| || |||| || || || || ||| ||
Db   631 TTCTTTCCAAAATTCCCTATCATTAATGCAACGCACCACTTATTTCCATCCCCGTGTGGGAG 690

Qy   1577 AGATCGCACCAGCCCATCACATTGACCGTTAACA---TTTCGGGCAAGCTGGTGGGGATTTTG 1633
          || | | |||| || || | | ||||| || || ||| |||| || || || ||
Db   691 AGCTTGCGCGCCGGCCATTCACGCCACTCTTCAAACGTATAATGACACGTTAGTTGAGTGTCT 750

Qy   1634 TCGTGACACACTTTGGGAACCACGAAGATGAGACCTCGACAGGAAACTGCAGGGCTATTGCTG 1693
          | |||||||| || ||| | || || || ||| || ||||| ||| || ||| || ||
Db   751 TTGTCTTTCCATACTGGCACAAAGAGAGCAAGATGAAGAAGAGATTACAAAAGTAATTACA 810

Qy   1694 TTTCAAAACTACTGAAAAAGTAGCTCTAATCAAGTGATATATTTCTGGGATTATATCACTTCAG 1753
          || ||||| | || || ||| | ||| ||| || |||||| | |||| |||| || || ||
Db   811 TGGCTAAATTGATGGAAATTCGACCCAGCTATTTTATTTAAGTTACTTGGTTGGTTG 870

Qy   1754 CACCTGGCTCCAGAGATTATCTACAGCTCACT--GAACATGGCAAATGTGAAGGATATCGA 1811
          || ||| ||| || || || ||| ||| || |||| || ||| |||||| || ||| |
Db   871 ACCGGGCGCGAAGGTAACTACAATACGCATGTTAGTGAACTTCTGGCCATGCATGACATTGA 930

```
Qy    1812 CAGCACTGATCATGACAGATGGTGTGAAATACATTATGTATCGGAGGCTCATCGTCAGGTTGGG 1871
           ||||||| ||||||||||| ||||||| |||||| || |||||| ||| ||| || |||:
Db     931 CCCTACTGACCGATAGGTGGTGTG-ATATATCTTGTATAAAGATTTGAGAGACAGGK 989

Qy    1872 T 1872
           |
Db     990 T 990
```

RESULT 14
LOCUS   AC114455
DEFINITION Rattus norvegicus clone CH230-249L20, *** SEQUENCING IN PROGRESS
            ***, 49 unordered pieces.
ACCESSION   AC114455
VERSION   AC114455.6 GI:21745639
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Norway rat.
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 162312)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,M., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,M., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 162312)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 162312)
  AUTHORS   Worley,K.C.
```

TITLE   Direct Submission
JOURNAL   Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          On Jul 14, 2002 this sequence version replaced gi:20467645.
COMMENT   -------------- Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
          Center project name: GSXI
          Center clone name: CH230-249L20
          -------------- Summary Statistics
          Sequencing vector: Plasmid;
          Chemistry: Dye-terminator Big Dye; version 0.990329
          Assembly program: Phrap; version 0.990329
          Consensus quality: 122635 bases at least Q40
          Consensus quality: 126261 bases at least Q30
          Consensus quality: 129045 bases at least Q20
          ----
          * NOTE: Estimated insert size may differ from sequence length
          * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 49 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *   1175 1174: contig of 1174 bp in length
          *   1275 1274: gap of unknown length
          *   2899 2899: contig of 1625 bp in length
          *   2900 2999: gap of unknown length
          *   3000 4263: contig of 1264 bp in length
          *   4264 4363: gap of unknown length
          *   4364 5667: contig of 1304 bp in length
          *   5668 5767: gap of unknown length
          *   5768 7230: contig of 1463 bp in length
          *   7231 7330: gap of unknown length
          *   7331 8913: contig of 1583 bp in length
          *   8914 9013: gap of unknown length
          *   9014 10114: contig of 1101 bp in length
          *   10115 10214: gap of unknown length
          *   10215 11240: contig of 1026 bp in length
          *   11241 11340: gap of unknown length
          *   11341 12994: contig of 1654 bp in length
          *   12995 13094: gap of unknown length
          *   13095 14158: contig of 1064 bp in length
          *   14159 14258: gap of unknown length
          *   14259 15405: contig of 1147 bp in length
          *   15406 15505: gap of unknown length
          *   15506 17366: contig of 1861 bp in length
          *   17367 17466: gap of unknown length
          *   17467 18850: contig of 1384 bp in length
          *   18851 18950: gap of unknown length
          *   18951 21635: contig of 2685 bp in length
          *   21636 21735: gap of unknown length
          *   21736 22866: contig of 1131 bp in length
          *   22867 22966: gap of unknown length
          *   22967 24541: contig of 1575 bp in length
          *   24542 24641: gap of unknown length
          *   24642 26761: contig of 2120 bp in length
          *   26762 26861: gap of unknown length
          *   26862 28689: contig of 1828 bp in length
          *   28690 28789: gap of unknown length
          *   28790 31195: contig of 2406 bp in length
          *   31196 31295: gap of unknown length
          *   31296 34187: contig of 2892 bp in length
          *   34188 34287: gap of unknown length
          *   34288 37140: contig of 2853 bp in length
          *   37141 37240: gap of unknown length
```

```
RESULT 15
Db    16227  TGGTACACATCGTTAAACCCAATTTGGACCTATCAGATGTCCAACAGAGTCATTCTGTCA 16286
Qy     475   TTAAGTGCCATAGCCACACTTGATCGTATTGGCACAGATGGTGA 518
             |||||||||||| || ||| ||| ||||| ||||||| ||
Db    16287  CTGAGTGCCATAGCTGTGTTCGACCGAATTAGCACAGGTAAGGA 16330
```

```
Db    16227
Qy     475   TTAAGTGCCATAGCCACACTTGATCGTATTGGCACAGATGGTGA 518
```

LOCUS   AC106576/c

```
LOCUS       AC106576        170472 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-198N10, *** SEQUENCING IN PROGRESS
            ***, 71 unordered pieces.
ACCESSION   AC106576
VERSION     AC106576.2  GI:21732007
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 170472)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Caroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,H., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,D., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,F., Zhou,J., Zorrilla,S., Nelson,D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170472)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 170472)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
```

FEATURES   Location/Qualifiers
     source   1..162312
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-249L20"

```
BASE COUNT    40948 a  33479 c  33666 g  42210 t  10709 others
ORIGIN
```

```
*  37241  38773: contig of 1533 bp in length
*  38774  41274: gap of unknown length
*  41275  42704: contig of 1330 bp in length
*  42705  42804: gap of unknown length
*  42805  44612: contig of 1808 bp in length
*  44613  44712: gap of unknown length
*  44713  47204: contig of 2492 bp in length
*  47205  47304: gap of unknown length
*  47305  49564: contig of 2260 bp in length
*  49565  49664: gap of unknown length
*  49665  52047: contig of 2383 bp in length
*  52048  52147: gap of unknown length
*  52148  55151: contig of 3004 bp in length
*  55152  55251: gap of unknown length
*  55252  58346: contig of 3095 bp in length
*  58347  58446: gap of unknown length
*  58447  61931: contig of 3485 bp in length
*  61932  62031: gap of unknown length
*  62032  65718: contig of 3687 bp in length
*  65719  65818: gap of unknown length
*  65819  69991: contig of 4173 bp in length
*  69992  70091: gap of unknown length
*  70092  73876: contig of 3785 bp in length
*  73877  73976: gap of unknown length
*  73977  77758: contig of 3782 bp in length
*  77759  77858: gap of unknown length
*  77859  80356: contig of 2498 bp in length
*  80357  80456: gap of unknown length
*  80456  82955: contig of 5627 bp in length
*  84522  84521: gap of unknown length
*  86084  86183: contig of 4491 bp in length
*  86184  90591: gap of unknown length
*  90592  90691: contig of 4408 bp in length
*  90692  94421: gap of unknown length
*  94422  94521: contig of 3730 bp in length
*  99012  99013: gap of unknown length
*  99013  99112: contig of 3788 bp in length
*  99113  102900: gap of unknown length
* 102901 103000: contig of 3277 bp in length
* 103001 106277: gap of unknown length
* 106278 106377: contig of 4976 bp in length
* 106378 111353: gap of unknown length
* 111354 111453: contig of 7341 bp in length
* 111454 118794: gap of unknown length
* 118795 118894: contig of 5287 bp in length
* 118895 124181: gap of unknown length
* 124182 124281: contig of 9323 bp in length
* 124282 133604: gap of unknown length
* 133605 133704: contig of 7324 bp in length
* 133705 141028: gap of unknown length
* 141029 141128: contig of 8475 bp in length
* 141129 149603: gap of unknown length
* 149604 149703: contig of 12609 bp in length
* 149704 162312: gap of unknown length
```

```
Best Local Similarity   78.7%;
Matches  129;  Conservative  0;  Mismatches  35;  Indels  0;  Gaps  0;
```

```
Qy     355  AGGTACCTCAGAATTTGGGGATTCATTTAGGACAGATTGTTCTTGTTGTTCTTACGGATA 414
            |||||||||| ||| ||||||||||||| |  ||||||| ||||||||||||| | ||||
Db   16167  AGGTACCTCAAAATCTGGGGATTCATTCTTCTTCTTGTTGTTCTTCTCCGAATA 16226
```

```
Qy     415  TGGTATACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACA 474
            |||||  |  ||  ||| |||| |||    ||  ||| |    ||  ||| ||  || ||
Db   16167
```

REFERENCE   AUTHORS   TITLE   JOURNAL

```
COMMENT     of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Jul 11, 2002 this sequence version replaced gi:18139100.
            ------------ Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            ------------ Project Information
              Center project name: GLJY
              Center clone name: CH230-198N10
            ------------ Summary Statistics
              Sequencing vector: Plasmid;
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 107919 bases at least Q40
              Consensus quality: 112493 bases at least Q30
              Consensus quality: 116174 bases at least Q20
            NOTE: Estimated insert size may differ from sequence length
              (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            NOTE: This is a 'working draft' sequence. It currently
              consists of 71 contigs. The true order of the pieces
              is not known and their order in this sequence record is
              arbitrary. Gaps between the contigs are represented as
              runs of N, but the exact sizes of the gaps are unknown.
              This record will be updated with the finished sequence
              as soon as it is available and the accession number will
              be preserved.
*
*        1    1338: contig of 1338 bp in length
*     1339    1438: gap of unknown length
*     1439    2513: contig of 1075 bp in length
*     2514    2613: gap of unknown length
*     2614    3930: contig of 1317 bp in length
*     3931    4030: gap of unknown length
*     4031    5174: contig of 1144 bp in length
*     5175    5274: gap of unknown length
*     5275    6681: contig of 1407 bp in length
*     6682    6781: gap of unknown length
*     6782    7973: contig of 1192 bp in length
*     7974    8073: gap of unknown length
*     8074    9242: contig of 1169 bp in length
*     9243    9342: gap of unknown length
*     9343   10738: contig of 1396 bp in length
*    10739   10838: gap of unknown length
*    10839   11895: contig of 1057 bp in length
*    11896   11995: gap of unknown length
*    11996   13088: contig of 1093 bp in length
*    13089   13188: gap of unknown length
*    13189   14286: contig of 1098 bp in length
*    14287   14386: gap of unknown length
*    14387   15530: contig of 1144 bp in length
*    15531   15630: gap of unknown length
*    15631   17385: contig of 1755 bp in length
*    17386   17485: gap of unknown length
*    17486   18802: contig of 1317 bp in length
*    18803   18902: gap of unknown length
*    18903   20384: contig of 1482 bp in length
*    20385   20484: gap of unknown length
*    20485   21552: contig of 1068 bp in length
*    21553   21652: gap of unknown length
*    21653   22803: contig of 1151 bp in length
*    22804   22903: gap of unknown length
*    22904   24096: contig of 1193 bp in length
*    24097   24196: gap of unknown length
*    24197   25575: contig of 1379 bp in length
*    25576   25675: gap of unknown length
*    25676   27490: contig of 1815 bp in length
*    27491   27590: gap of unknown length
*    27591   28839: contig of 1249 bp in length
*    28840   28939: gap of unknown length
*    28940   30343: contig of 1404 bp in length
*    30344   30443: gap of unknown length
*    30444   31581: contig of 1138 bp in length
*    31582   31681: gap of unknown length
*    31682   33179: contig of 1498 bp in length
*    33180   33279: gap of unknown length
*    33280   34298: contig of 1019 bp in length
*    34299   34398: gap of unknown length
*    34399   36062: contig of 1664 bp in length
*    36063   36162: gap of unknown length
*    36163   37416: contig of 1254 bp in length
*    37417   37516: gap of unknown length
*    37517   39068: contig of 1552 bp in length
*    39069   39168: gap of unknown length
*    39169   40889: contig of 1721 bp in length
*    40890   40989: gap of unknown length
*    40990   43123: contig of 2134 bp in length
*    43124   43223: gap of unknown length
*    43224   44769: contig of 1546 bp in length
*    44770   44869: gap of unknown length
*    44870   46445: contig of 1576 bp in length
*    46446   46545: gap of unknown length
*    46546   47649: contig of 1104 bp in length
*    47650   47749: gap of unknown length
*    47750   49233: contig of 1484 bp in length
*    49234   49333: gap of unknown length
*    49334   52066: contig of 2733 bp in length
*    52067   52166: gap of unknown length
*    52167   53437: contig of 1271 bp in length
*    53438   53537: gap of unknown length
*    53538   55843: contig of 2306 bp in length
*    55844   55943: gap of unknown length
*    55944   57376: contig of 1433 bp in length
*    57377   57476: gap of unknown length
*    57477   59067: contig of 1591 bp in length
*    59068   59167: gap of unknown length
*    59168   60899: contig of 1732 bp in length
*    60900   60999: gap of unknown length
*    61000   62740: contig of 1741 bp in length
*    62741   62840: gap of unknown length
*    62841   64298: contig of 1458 bp in length
*    64299   64398: gap of unknown length
*    64399   66078: contig of 1680 bp in length
*    66079   66178: gap of unknown length
*    66179   68380: contig of 2202 bp in length
*    68381   68480: gap of unknown length
*    68481   70438: contig of 1958 bp in length
*    70439   70538: gap of unknown length
*    70539   72478: contig of 1940 bp in length
*    72479   72578: gap of unknown length
*    72579   75318: contig of 2740 bp in length
*    75319   75418: gap of unknown length
*    75419   78197: contig of 2779 bp in length
*    78198   78297: gap of unknown length
*    78298   80521: contig of 2224 bp in length
*    80522   80621: gap of unknown length
*    80622   83881: contig of 3260 bp in length
*    83882   83981: gap of unknown length
*    83982   86825: contig of 2844 bp in length
*    86826   86925: gap of unknown length
*    86926   90212: contig of 3287 bp in length
*    90213   90312: gap of unknown length
*    90313   93565: contig of 3253 bp in length
*    93566   93665: gap of unknown length
*    93666   96054: contig of 2389 bp in length
```

```
Query Match
Best Local Similarity  78.2%;  Conservative  3.9%;  Score 81.6;  DB 2;  Length 170472;
Matches 111;  Pred. No. 2.9e-11;  Mismatches 29;  Indels 2;  Gaps 1;

Qy   1882 ATCTCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAGATCCCTGAT 1941
          || ||| |||||||||||||| ||| ||  ||  |||||| |||||| |||||||| ||
Db   2512 ATTTTCTCATGCTGAACTGAGTGAGTGACTCTGAGAT--CAGATGCCAAATTCAGGATCCCAGAT 2455

Qy   1942 GACCCCACTAATTATAGAGACAACCAGAAAAGTGGTCATAGACCACCAGAGAGAAGTTTCTGAG 2001
```

```
Db  2454  GATCCCTTCAATTACAGAGACAACCAGAGAGTCGTTCATAGACCACAGAGGAGTTCCCAAG  2395
           || |||  |||||  |||||||||||  |||  ||||||||||||||||||||  |||| | ||
Qy  2002  AAAATTCATTTTTAATCCCAGAT  2023
           ||  |||||||||  ||  ||  ||| |
Db  2394  AACATTCATTTCAACCCGAGGT  2373
```

Search completed: January 7, 2003, 19:04:12
Job time : 9564 secs

ALIGNMENTS

```
RESULT 1
Q9H720
ID   Q9H720       PRELIMINARY;    PRT;    699 AA.
AC   Q9H720;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CDNA: FLJ21511 fis, clone COL05748.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON;
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA   Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK025164; BAB15080.1; -.
SQ   SEQUENCE   699 AA;  78565 MW;  2F839563189A0523 CRC64;
```

```
  Query Match              99.8%;  Score 3713;  DB 4;  Length 699;
  Best Local Similarity    99.9%;  Pred. No. 3.6e-263;
  Matches  698;  Conservative   1;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEGFSIAFLSPIFLTITPF 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEGFSIAFLSPIFLTITPF 60

Qy    61 WKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120
         |||||||||||||||||||||||||||||+:+|||||||||||||||||||||||||||||
Db    61 WKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120

Qy   121 LRIWGFILGQIVLVVVLRIWYTSLNPIWSYQMSNKVILTLSAIATLDRIGTDGDCSKPEEK 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 LRIWGFILGQIVLVVVLRIWYTSLNPIWSYQMSNKVILTLSAIATLDRIGTDGDCSKPEEK 180

Qy   181 KTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGA 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 KTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGA 240

Qy   241 VLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTASMWPQ 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 VLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTASMWPQ 300

Qy   301 TLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLII 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLII 360

Qy   361 GLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYERKLG 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYERKLG 420

Qy   421 KVAPTKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFITILESDASKPYMGNNDLT 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 KVAPTKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFITILESDASKPYMGNNDLT 480

Qy   481 MWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAITLTVNISGK 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 MWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAITLTVNISGK 540

Qy   541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNV 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNV 600

Qy   601 KDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 KDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660

Qy   661 DHREVSEKIHFNPRFGSYKEGHNYENNHHFHMNTPKYFL 699
         ||||||||||||||||||||||||||||:|:||||||||
Db   661 DHREVSEKIHFNPRFGSYKEGHNYENNHNFHMNTPKYFL 699
```

THIS PAGE BLANK (USPTO)

RESULT 1
AK025164
LOCUS   AK025164   2486 bp   mRNA   linear   PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21511 fis, clone COL05748.
ACCESSION   AK025164
VERSION   AK025164.1 GI:10437625
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL05748.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo..
REFERENCE   1 (sites)
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.

  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2486)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES   Location/Qualifiers
     source   1..2486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="COL05748"
                     /tissue_type="colon"
                     /clone_lib="COL"
                     /note="cloning vector pME18SFL3"
     CDS   183..2282
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15080.1"
                     /db_xref="GI:10437626"
                     /translation="MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEG
                     FSIAFLSPIFLTITPFWKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSS
                     SLIVQAVTWWSGSHLQRYLRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVILTLS
                     AIATLDRIGTDGDCSKPEEKKTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSL
                     VSRWAVSGHPHPGPDPNPFGGAVLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLL
                     YLHTWAAAVSGCVFAIFTASMWPQTLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCT
                     AFKFVPGGVYARERSDVLLGTMMLIIGLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKY
                     MKLFLWLLVGVGLLGLGLRHKAYERKLGKVAPTKEVSAAIWPFRFGYDNEGWSSLERS
                     AHLLNETGADFITILESDASKPYMGNNDLTMWLGEKLGFYTDFGPSTRYHTWGIMALS
                     RYPIVKSEHHLLPSPEGEIAPAITLTVNISGKLVDFVVTHFGNHEDDLDRKLQAIAVS
                     KLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDHDRWCEYIMYRGLIRLG
                     YARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVIDHREVSEKIHFNPRFGSYKEGH
                     NYENNHNFHMNTPKYFL".
BASE COUNT   692 a   516 c   606 g   672 t
ORIGIN

     Query Match   99.8%;   Score 2098.2;   DB 9;   Length 2486;
     Best Local Similarity   99.9%;   Pred. No. 0;
     Matches 2100;   Conservative   .0;   Mismatches   3;   Indels   0;   Gaps   0;

Qy   1 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   183 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 242

Qy   61 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 302

Qy   121 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||!||||||||
Db   303 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 362

Qy   181 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   363 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 422

Qy   241 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   423 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 482

Qy   301 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCGGGAAGTCATTTGCAAAGGTAC 360
           ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||

```
Db    483 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCAGGAAGTCATTTGCAAAGGTAC 542

Qy    361 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    543 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 602

Qy    421 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 662

Qy    481 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    663 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 722

Qy    541 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    723 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGCT 782

Qy    601 GCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    783 GCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTCC 842

Qy    661 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCATTGGAGGTGCA 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    843 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCATTGGAGGTGCA 902

Qy    721 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGTACT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    903 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGTACT 962

Qy    781 GGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    963 GGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 1022

Qy    841 TGGGCAGCTGCTGTGTCTGGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1023 TGGGCAGCTGCTGTGTCTGGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAA 1082

Qy    901 ACACTTGGACACCTTATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1083 ACACTTGGACACCTTATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATG 1142

Qy    961 ATATTTTATCTTCTAGAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1143 ATATTTTATCTTCTAGAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCA 1202

Qy   1021 GGAGGTGTCTACGCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1203 GGAGGTGTCTACGCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATC 1262

Qy   1081 GGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGACTTGCTTGTTCAAACAAAAAAC 1140
          ||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1263 GGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGATTTGCTTCTTCAAACAAAAAAC 1322

Qy   1141 AGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1323 AGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTT 1382

Qy   1201 GTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1383 GTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGC 1442

Qy   1261 AAAGTGGCACCAACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGAC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1443 AAAGTGGCACCAACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGAC 1502

Qy   1321 AATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1503 AATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGAT 1562

Qy   1381 TTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1563 TTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACC 1622
```

THIS PAGE BLANK (USPTO)

```
Qy   1441 ATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCAC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1623 ATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCAC 1682

Qy   1501 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1683 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTT 1742

Qy   1561 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1743 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAG 1802

Qy   1621 CTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1803 CTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTG 1862

Qy   1681 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1863 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1922

Qy   1741 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1923 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1800

Qy   1801 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 1982
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1983 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 1860

Qy   1861 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATG 2042
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2043 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATG 1920

Qy   1921 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 2102
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2103 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 1980

Qy   1981 GACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAA 2162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2163 GACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAA 2040

Qy   2041 GGACACAATTATGAAAACAACCATCATTTTCATATGAATACTCCCAAATACTTTTTATGA 2222
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   2223 GGACACAATTATGAAAACAACCATAATTTTCATATGAATACTCCCAAATACTTTTTATGA 2100

Qy   2101 AAC 2103
          |||
Db   2283 AAC 2285
```

THIS PAGE BLANK (USPTO)